Reconstructing the History of Maize Streak Virus Strain A Dispersal To Reveal Diversification Hot Spots and Its Origin in Southern Africa^{∇†}

Adérito L. Monjane,¹ Gordon W. Harkins,² Darren P. Martin,^{3,4*} Philippe Lemey,⁵ Pierre Lefevre,^{3,6} Dionne N. Shepherd,¹ Sunday Oluwafemi,⁷ Michelo Simuyandi,⁸ Innocent Zinga,⁹ Ephrem K. Komba,⁹ Didier P. Lakoutene,⁹ Noella Mandakombo,⁹ Joseph Mboukoulida,⁹ Silla Semballa,⁹ Appolinaire Tagne,¹⁰ Fidèle Tiendrébégo,¹¹ Julia B. Erdmann,^{1,12} Tania van Antwerpen,¹³ Betty E. Owor,^{1,14} Bradley Flett,¹⁵ Moses Ramusi,¹⁵ Oliver P. Windram,¹⁶ Rizwan Syed,¹ Jean-Michel Lett,⁶ Rob W. Briddon,¹⁷ Peter G. Markham,¹⁸ Edward P. Rybicki,^{1,3} and Arvind Varsani^{19,20,21}

Department of Molecular and Cell Biology, University of Cape Town, Rondebosch, 7701, Cape Town, South Africa¹; South African National Bioinformatics Institute, University of the Western Cape, Cape Town, South Africa²; Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Observatory, 7925, Cape Town, South Africa³; Centre for High-Performance Computing, Rosebank, Cape Town, South Africa⁴; Department of Microbiology and Immunology, Rega Institute, K.U. Leuven, Leuven, Belgium⁵; CIRAD, UMR 53 PVBMT CIRAD-Université de la Réunion, Pôle de Protection des Plantes, 97410, Saint Pierre, La Réunion, France⁶; Department of Crop Production, Soil and Environmental Management, Bowen University, Iwo, Osun State, P.M.B. 284, Nigeria⁷; P.O. Box RW 50138, Lusaka, Zambia⁸; LASBAD Laboratory, Faculty of Sciences, University of Bangui, BP 908 Bangui, Central African Republic⁹; Cereals Research Program, Institute of Agricultural Research for Development, Box 2067 Messa, Yaounde, Cameroon¹⁰; Centre de Recherche en Sciences Biologiques Alimentaires et Nutritionnelles (CRSBAN), UFR/SVT Université de Ouagadougou, 03 BP 7131 Ouagadougou 03, Burkina Faso¹¹; Institute of Biology, Department of Molecular Biology and Plant Virology, University of Stuttgart, Pfaffenwaldring 57, D-70550 Stuttgart, Germany¹²; South African Sugarcane Research Institute, Mount Edgecombe, KwaZulu Natal, South Africa¹³; Department of Plant Sciences, University of Cambridge, Cambridge, United Kingdom, CB2 3EA¹⁴; Crop Protection, ARC-Grain Crops Institute, Potchefstroom 2520, South Africa¹⁵; Warwick HRI Biology Centre, University of Warwick, Wellesbourne, CV35 9EF, England¹⁶; National Institute for Biotechnology and Genetic Engineering, Jhang Road, P.O. Box 577, Faisalabad, Pakistan¹⁷; Department of Disease and Stress Biology, John Innes Centre, Norwich NR4 7UH, United Kingdom¹⁸; Biomolecular Interaction Centre, University of Canterbury, Private Bag 4800, Christchurch, 8140, New Zealand¹⁹; School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch, 8140, New Zealand²⁰; and Electron Microscope Unit, University of Cape Town, Rondebosch, 7701, Cape Town, South Africa²¹

Received 30 March 2011/Accepted 21 June 2011

Maize streak virus strain A (MSV-A), the causal agent of maize streak disease, is today one of the most serious biotic threats to African food security. Determining where MSV-A originated and how it spread transcontinentally could yield valuable insights into its historical emergence as a crop pathogen. Similarly, determining where the major extant MSV-A lineages arose could identify geographical hot spots of MSV evolution. Here, we use model-based phylogeographic analyses of 353 fully sequenced MSV-A isolates to reconstruct a plausible history of MSV-A movements over the past 150 years. We show that since the probable emergence of MSV-A in southern Africa around 1863, the virus spread transcontinentally at an average rate of 32.5 km/year (95% highest probability density interval, 15.6 to 51.6 km/year). Using distinctive patterns of nucleotide variation caused by 20 unique intra-MSV-A recombination events, we tentatively classified the MSV-A isolates into 24 easily discernible lineages. Despite many of these lineages displaying distinct geographical distributions, it is apparent that almost all have emerged within the past 4 decades from either southern or east-central Africa. Collectively, our results suggest that regular analysis of MSV-A genomes within these diversification hot spots could be used to monitor the emergence of future MSV-A lineages that could affect maize cultivation in Africa.

Maize streak disease (MSD) is one of the most serious biotic constraints to maize production in sub-Saharan Africa. It has a complex and poorly understood epidemiology, with periodic

outbreaks occurring that devastate the maize yields of small- and medium-scale farmers (42, 61). Considering the overwhelming contribution of maize to the total daily caloric intake of Africans—equaled only by wheat—and the generally low per-hectare maize yields throughout the continent, MSD persistently threatens the food security and economic empowerment of the more than 100 million impoverished Africans who are dependent on subsistence farming (15, 42).

Maize streak virus (MSV) (genus *Mastrevirus*, family *Geminiviridae*), the causal agent of MSD, is distributed throughout the African continent and surrounding islands (61). As is typical of all mastreviruses, its circular, ~2.7-kb monopartite sin-

* Corresponding author. Mailing address: Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Private Bag X3, Rondebosch 7701, Cape Town, South Africa. Phone: 27 21 406 6366. Fax: 27 21 689 1528. E-mail: darrenpatrickmartin@gmail.com.

† Supplemental material for this article may be found at <http://jvi.asm.org/>.

[∇] Published ahead of print on 29 June 2011.

gle-stranded DNA (ssDNA) genome encodes only four proteins. Bidirectional transcription from a long intergenic region (LIR) results in the virion sense expression of a movement protein (MP) and a coat protein (CP) and the complementary-sense expression of the replication-associated proteins, Rep and RepA. Whereas MP and CP are involved in virus movement and encapsidation (6, 32, 72), Rep is an essential initiator of virus replication, and RepA is a regulator of host and viral gene transcription (9, 22, 23, 33, 73). Due to genome size restrictions, MSV usurps host DNA replication and double-stranded DNA break repair proteins to replicate its genome via, respectively, rolling-circle (58) and recombination-dependent (12, 25) mechanisms.

Taxonomically, viruses within mastrevirus species (i.e., groups of mastreviruses sharing greater than 75% genome-wide sequence similarity) have been subcategorized into strains containing isolates sharing >91% genome-wide sequence similarity and subtypes containing isolates sharing >98% similarity (13, 39). To date, 11 distinct MSV strains (classified as MSV-A through -K [39, 59, 69, 71]), have been identified. Whereas the majority of these are apparently adapted to infecting wild grass species in the genera *Digitaria*, *Urochloa*, and *Setaria* (57, 69), only MSV-A isolates are adapted to infecting maize (39). The contribution of recombination to the genetic diversification of MSV, as with other geminiviruses (34, 49, 52), is evident in the observation that nearly all strains apart from MSV-G, -I, and -E are interstrain recombinants. In addition, MSV-F and -H and some MSV-B variants are the recombinant progeny of MSV and other African streak mastreviruses, such as *Panicum streak virus* (PanSV) and *Saccharum streak virus* (31, 39).

Maize was first introduced to West Africa by the Portuguese in the early 1500s and to southern Africa by the Dutch in the mid-1600s (24, 44), and a recombination event between ancestral MSV-B and MSV-G/F variants is credited with having generated a maize-adapted MSV-A prototype (54), possibly within 20 years of the first credible reports of MSD in southern Africa during the 1870s (17, 21). The estimated timing of this recombination event, the observation that it involved the exchange of an experimentally demonstrated pathogenicity-determining *mp* and *cp* gene module (40), the fact that it is detectable in all known MSV-A variants (69), and the observation that MSV-A is the only MSV strain that is associated with MSD (7, 39, 47) have prompted speculation that this recombination event was key to the emergence of MSV-A as a viral pathogen of maize (21).

Since its initial emergence and subsequent spread throughout the continent, five distinct MSV-A subtypes—MSV-A₁, -A₂, -A₃, -A₄, and -A₆—have evolved, and today, each has a different well-defined geographical range within sub-Saharan Africa (69). Although it is evident that over the last century the movement of some MSV-A subtypes across Africa has been far less constrained than the movements of either other MSV strains (69) or related African streak virus species (70), it is also apparent that the different MSV-A subtypes are not equally mobile. For example, whereas the MSV-A₁ subtype is found throughout Africa, the MSV-A₂, -A₃, -A₄, and -A₆ subtypes have been found only in West Africa, East Africa, southern Africa, and the Indian Ocean island of La Réunion, respectively (39, 50, 69).

Emphasizing the role of recombination between different

MSV-A lineages (known as intrastrain recombination) in the ongoing diversification of the MSV-A subtypes, Owor et al. (47) further classified a large number of randomly sampled MSV-A₁ isolates from Uganda into recombinant lineage groupings (called “haplotypes” in that publication) based on the patterns of intrastrain recombination events detectable within their genomes. While this study revealed that in Uganda in 2006 there existed a single predominant and widely distributed MSV-A₁ recombinant lineage (i.e., a group of closely related viruses all carrying evidence of the same recombination events), it also revealed geographical differences in the population-wide frequencies of various other MSV-A₁ lineages.

Detailed information on the spatial dynamics of different MSV-A lineages throughout Africa is of great potential interest to plant pathologists, epidemiologists, maize breeders, and farmers, who are intent on mitigating the effects of MSD on African maize production. Of particular relevance in this regard is the recent development of sophisticated DNA sequence analysis techniques that, in conjunction with information on sampling dates and geographical coordinates, can provide statistically rigorous inferences of changes in the spatial distributions of viral pathogens through time. By identifying the geographical origins of emergent epidemic variants, their rates of movement, and their migration pathways (16, 30, 35), such techniques could certainly help elucidate the various factors that impact the epidemiological dynamics of MSD (37, 63, 66).

Among these are vector- and human-mediated factors that strongly influence the dynamics of MSD epidemics. For example, among the nine species of *Cicadulina* leafhoppers known to transmit MSV (3, 10, 14, 45, 46, 54, 55, 64), *Cicadulina mbila* and *Cicadulina storeyi* are more efficient MSV transmitters than other species (1, 45). The species compositions, sizes, and distributions of leafhopper populations are known to vary across Africa and are strongly influenced by the seasonal and geographical availability of the plants that leafhoppers feed on (5). Besides natural variations in leafhopper migration dynamics (2, 56), the cultivation of maize “green corridors” throughout the continent and possible human-mediated movements of infected maize or MSV-containing leafhoppers have potentially also impacted the epidemiological dynamics of MSD by facilitating the unnaturally long-range dispersal of MSV-A.

Here, we use geospatial sampling coordinates, full MSV genome sequences, and detailed recombination analyses to map the distributions of MSV-A recombinant lineages throughout most of Africa’s main maize-growing areas during the period 2000 to 2008. In addition to identifying many novel MSV-A₁ and MSV-A₄ recombinant lineages and presenting a detailed up-to-date snapshot of the current spatial distribution of MSV-A diversity on a continental scale, we use sophisticated new Bayesian phylogeographic techniques to reconstruct the spatial and temporal dynamics of MSV-A distributions over the past century. Specifically, we infer the geographical locations of key ancestral MSV-A genotypes and determine historical MSV-A dispersal pathways and movement rates to reconstruct a plausible time line for the spatial dissemination of MSV-A since its emergence in the mid-19th century.

MATERIALS AND METHODS

Virus isolates. A total of 193 MSV-A isolates were obtained from maize, wild grasses, and sugarcane plants displaying characteristic streak virus symptoms

sampled in West, East, Central, and southern Africa and the Indian Ocean island of La Réunion between 1988 and 2008. The geographical locations, sampling dates, and hosts of the 193 new field isolates are listed in Table S1 in the supplemental material. In addition to 160 publicly available MSV-A full-genome sequences that were included in all of our sequence analyses, we included restriction fragment length polymorphism (RFLP)-based haplotyping data from another 310 Ugandan isolates (47) during our assessment of the continent-wide geographical ranges of MSV-A recombinant lineages (see Table S1 in the supplemental material).

Genome cloning and sequencing. Full-length viral genomes were amplified from total plant DNA using phi29 DNA polymerase (TempliPhi; GE Healthcare) as described by Owor et al. (48) and Shepherd et al. (60). Amplified genome concatemers were digested with either KpnI or BamHI (see Table S1 in the supplemental material for details) to yield linearized viral genomes (~2.7 kb), which were ligated into analogously linearized pGEMZf+ (Promega Biotech). Using primers described previously (48), both strands of each full-length genome were subjected to DNA sequencing (Macrogen Inc., South Korea). The sequences were edited and assembled using DNAMAN (version 5.2.9; Lynnon Biosoft, Canada) and Mega 4 (67).

Phylogenetic and recombination analyses. Using POA v.2 (19), the 193 new MSV-A genome sequences were aligned with 178 publicly available African streak virus full-genome sequences. These comprised 160 other full MSV-A genomes, 1 representative genome from each of the 10 other known MSV strains (strains B through K), and 1 sequence from each of the 8 different non-MSV African streak virus species (see Table S1 in the supplemental material). Multiple-sequence alignments generated using the Clustal W-based (68) alignment function in Mega 4 (67) were edited by visual inspection. Mega 4 was also used to calculate average pairwise Hamming distances (maximum composite likelihood method with pairwise exclusion of gaps) between the various MSV-A genomes.

Maximum-likelihood (ML) phylogenetic trees were constructed using PHYLML v.1 (20) with the GTR+I+G₄ nucleotide substitution model (selected as the most appropriate by RDP3 [43]).

Recombination was analyzed using the RDP (38), GENECONV (49), RECSAN (41), MAXCHI (62), CHIMAERA (51), SISCAN (18), and 3SEQ (4) recombination detection methods implemented in RDP3 (43). Only potential recombination events detected by two or more of these methods, together with phylogenetic evidence of recombination, were considered robust evidence of recombination.

We first analyzed the multiple-sequence alignment for evidence of interspecies and interstrain recombination. When such evidence was found, the sequence tracts that had introgressed into the MSV-A isolates were removed from the alignment and the resulting gaps were treated as missing data. Non-MSV-A full genomes were subsequently removed, and the remaining alignment is here referred to as the “full-genome” (FG) data set (see data set S1 in the supplemental material). This data set was free of interspecies and interstrain recombination but retained evidence of intrastrain recombination events. Further recombination analysis of the FG data set was carried out to identify these intrastrain recombination events.

Following the characterization of intrastrain recombination events, we identified a 1,042-nucleotide (nt)-long fragment of the FG alignment (roughly corresponding to the movement and coat protein genes) that was mostly free of recombination breakpoints. We produced an “MPCP data set” from this recombination-free portion of the FG data set by deleting the remainder of the alignment and removing 20 sequences with detectable recombination breakpoints within this region. This 333-sequence (1,042 nt in length) MPCP data set (alignment available from the authors upon request) was found to be free of detectable recombination by using the PHI test (with default settings [8]; while more sensitive to the presence of recombination than the other tests we applied, this test is incapable of identifying recombinant sequences and recombination breakpoints) and was used as our “recombination-free” data set for subsequent phylogeographic analyses. To produce a phylogenetic tree that was not strongly influenced by recombination yet contained representative sequences from as many currently sequenced full genomes as possible, we produced a second recombination-free data set that included all but one of the sequenced MSV genomes (i.e., 352 of the sequences) but contained only 635 nt of these sequences corresponding to the CP gene. This recombination-free data set, called the “CP data set” (alignment available from the authors upon request) was used, together with the FG data set, to classify the sequenced MSV genomes into different recombinant lineages.

Phylogeographic analyses. A phylogeographic analysis of the FG data set and the MPCP data set was conducted using a Bayesian Markov chain Monte Carlo approach that used probabilistic models of ancestral virus movement to provide

plausible explanations for the current geographic distributions of the sequences in the two data sets (35, 36). Implemented in the program BEAST (11), the methods we applied incorporated geographical information as distance-informed priors on the rates at which ancestral viruses have moved across the African continent, i.e., prior expectations were that viruses would have a greater probability of moving between adjacent countries than between countries on opposite sides of the continent. We either restricted spatial coordinates to discrete locations (an approach that involves grouping sequences based on, for example, their countries of origin so that all sequences in a given country share the same discrete geographical coordinate [35]) or approximated them as more spatially realistic virus dispersal patterns by explicitly modeling the diffusion process on a continuous landscape (an approach in which each input sequence and all inferred ancestral sequences can have their own unique geographical coordinates). The continuous models can be further divided into those that employ strict Brownian diffusion to model spatial movement over the entire phylogeny so that the same rate applies to all branches at all times and between any two locations and those that relax this assumption to allow for heterogeneity in diffusion rates (36). The relaxed Brownian diffusion models have been shown to result in improved statistical efficiency in spatial reconstructions of overdispersed (i.e., with a variance greater than the mean) random walks (35). Both data sets were analyzed under strict and relaxed spatial diffusion models, and whereas the FG data set runs converged to stationary irrespective of model choice, the MPCP data set runs failed to do so under the strict model. Collectively, these analyses enabled inference of the timing, direction, and rates of viral movements, along with the most probable geographical coordinates of ancestral MSV-A sequences (such as the most recent common ancestor [MRCA] of all the MSV-A sequences or the MRCAs of isolates belonging to the various recombination lineages).

To determine the approximate statistical support for the inferred movement pathways, we implemented a Bayes factor (BF) test, where evidence of virus movements between different locations yielding a BF of <3 is not well supported, those locations yielding a BF of >3 are indicative of substantial support, and BFs of >10 and >100 are indicative of strong and decisive support (26, 65).

We analyzed the FG and MPCP data sets with both the discrete and continuous phylogeographic models implemented in BEAST under a relaxed-clock model assuming the Hasegawa-Kishino and Yano substitution model with four gamma-distributed substitution rate categories (HKY+G₄), and the flexible nonparametric Bayesian skyline plot (BSP) demographic model (selected for MSV as the best-fit clock) (the substitution and demographic models have been described by Harkins et al. [21].) Replicate runs of the Markov chain varying in length between 2×10^8 and 8×10^8 steps were performed with BEAST and inspected for convergence using Tracer v1.6 (53). When similar results were obtained, the independent replicate runs were combined using LogCombiner, available as part of the BEAST package. The estimated sample sizes (ESS) of the combined files were >200, indicating sufficient mixing of the Markov chain and parameter sampling.

Whereas the discrete model sequences were grouped according to their countries of origin, the continuous model made use of the individual geographical coordinates of every analyzed sequence. The results of analyses indicating the most plausible set of MSV-A movements across Africa during the past century were formatted for viewing as animations within Google Earth using the tools available from http://beast.bio.ed.ac.uk/Google_Earth.

Nucleotide sequence accession numbers. The annotated genome sequences were submitted to GenBank under the accession numbers indicated in Table S1 in the supplemental material (accession numbers HQ693281 to HQ693474).

RESULTS AND DISCUSSION

Characterization of MSV-A recombinants. For purposes of classifying the available MSV-A genomes into recombinant lineage groupings, it was necessary to first characterize the intrastrain recombination patterns present within these genomes. The 193 MSV-A genomes determined here were analyzed together with 160 previously described MSV-A genomes, 10 representative genomes from non-MSV-A strains, and 8 from other African streak virus species obtained from public sequence databases (see data set S1 in the supplemental material). Whereas we found no evidence of any of the 353 MSV-A isolates having acquired sequences through interspecies recombination with other African streak viruses, the recombination analysis revealed that 98 of the MSV-A isolates

collectively carried evidence of four unique interstrain recombination events. Three of these have been described previously. They involved 88 MSV-A₄ isolates carrying MSV-B-like sequences at the 3' ends of their replication-associated protein genes (*rep*) (39), two MSV-A₁ isolates carrying MSV-G-like sequences at the 3' ends of their *rep* genes (69 and data not shown), and seven MSV-A₄ isolates carrying MSV-B-like sequences within their *mp* genes (data not shown) (71). The fourth recombination event, revealed for the first time by the analysis conducted here, involved the introduction of ~69 nt of MSV-G strain-like sequence (data not shown) within the 3' end of the *rep* gene of MSV-A (BF-Gol-BF1-2008).

The absence of any evidence of interspecies recombination within the analyzed MSV-A sequences provides additional support to the notion that interspecies recombination events involving MSV-A isolates either occur only very rarely or yield genomes that are so defective that the action of purifying selection, operating at any point from the cellular to the ecosystem scale, ensures that they are maintained only at very low frequencies in natural populations (69).

For subsequent analyses, the non-MSV-A sequences were removed from the alignment. This step ensured that phylogenetic noise caused by these recombination events did not obscure more subtle signals of intra MSV-A recombination events. We refer to the exclusively MSV-A alignment thus obtained as our FG data set.

When the FG data set was rescreened for intrastrain recombination, we identified a total of 16 well-supported unique recombination events, all within MSV-A₁ and -A₄ sequences. To this list we added events 2^s, 3^s, and 9^s, which were identified previously by Owor et al. (47) (Table 1). Although 11 of the 16 recombination events detected here were also identified by Owor et al. (47) and Varsani et al. (69), our larger sequence data set afforded additional clarity to the interpretation of these events. A subset of MSV-A₁ isolates previously identified as lacking evidence of intra-MSV-A recombination (47) are shown here to have a short (401-nt) segment exchanged between MSV-A₁ viruses (event 12 in Table S1 in the supplemental material).

Classification of MSV-A genomes. Both the recombinant-lineage naming convention and the recombinant-lineage names designated by Owor et al. (47) (where the authors referred to them as "haplotypes") were maintained here, except that the "Ug" prefix was removed from all the names. For example, MSV-A₁UgI is given here as MSV-A₁I, or recombinant lineage I within the MSV-A₁ subtype. Whereas recombinant-lineage names were previously given only for subtype MSV-A₁, here, we assigned names to all isolates in the other MSV-A subtypes. The virus genomes within the MSV-A₁, -A₂, -A₃, -A₄, and -A₆ subtypes that were not detectably intra-MSV-A recombinants were assigned to lineages MSV-A₁XV, MSV-A₂I, MSV-A₃II, MSV-A₄V, and MSV-A₆I, respectively (diagrammatic depictions of these are shown in Fig. 1).

The MSV-A viruses were thus classified into 19 recombinant lineages, 14 of which contained predominantly MSV-A₁-like sequences. The number of viral genomes assigned to individual recombinant lineages differed quite widely, with MSV-A₁IX, for example, being represented by a single sampled genome and MSV-A₁V being represented by 68 widely distributed isolates sampled between Central Africa and Mozambique.

Using an ML approach, phylogenetic trees of MSV-A isolates sampled between 1979 and 2008 were constructed using data sets composed of either 353 fully sequenced MSV genomes (the FG data set) (Fig. 1) or 352 CP gene-containing genome fragments that were determined to be mostly recombination free (the CP data set) (see Fig. S1 in the supplemental material). The phylogenetic trees constructed from these two data sets shared similar topologies, so that the individual MSV-A subtype clades and many of the individual recombinant-lineage clades were well resolved in both trees (Fig. 1; see Fig. S1 in the supplemental material). However, bootstrap support for these groupings was generally poorer in the CP tree than in the FG tree, possibly owing to the shorter length of sequences (635 versus 2,689 nt) used to construct the former tree.

In general, viruses represented within the trees displayed a high degree of geographical clustering. At the subtype level, the geographical ranges of the different groups (MSV-A₁ to -A₄, and -A₆) were identical to those reported previously (39, 50, 69). Therefore, whereas the MSV-A₁ subtype had a continent-wide distribution, the MSV-A₂, -A₃, -A₄, and -A₆ subtypes were found only in West Africa, East Africa, southern Africa, and the island of La Réunion, respectively.

At 2.6%, the overall mean genetic distance (using the maximum-likelihood method implemented in Mega 4 with the exclusion of gaps as missing data) between the various MSV-A genomes was short. However, notable regional differences in the mean genetic distances (as calculated using Mega 4) between the sampled viruses was such that the mean diversities of the analyzed MSV-A genomes from West Africa, East Africa, and southern Africa were 1%, 1.8%, and 2.6%, respectively.

Mapping the current geographical distribution of MSV-A recombinant lineages. Based on the evidence for geographical clustering within the FG and CP phylogenetic trees, we sought to define the current geographical distributions and relative prevalences in different parts of Africa of the various identified MSV-A recombinant lineages using viruses sampled between the years 2000 and 2008 (collectively belonging to 13 of the classified lineages). While determining the continent-wide distributions of these lineages, an additional 310 Ugandan RFLP haplotyped isolates (47) were also included in the analysis (see Table S1 in the supplemental material).

The recombinant lineages differed markedly with respect to their prevalences in different parts of Africa and in the extents of their geographical ranges (Fig. 2). These differences provide some clues about the potential geographical origins of the different lineages and the dynamics of their recent movements. For example, while MSV-A₁III accounts for >60% of Ugandan MSV isolates sampled between 2000 and 2008 (zones 3 to 8 in Fig. 2D), it is found in the adjacent western and central African regions (zones 1 and 2 in Fig. 2D) at only 2.3% and 9.1% of sampled MSV-A infections, a pattern suggesting that MSV-A₁III may be in the process of diffusing westward from Uganda. Conversely, the relative frequency of sampled MSV-A₁VII infections in these same regions suggests that this lineage may be in the process of diffusing eastward from West Africa (i.e., from zone 1 toward zone 8 in Fig. 2H).

Although low population densities may account for the detection of some lineages only within relatively small regions of Africa, other factors may be responsible for the minimal de-

TABLE 1. Intra-MSV-A recombination events detected in 353 fully sequenced MSV-A sequences

Event	Breakpoint position ^a		Recombinant(s)	Minor parent ^c	Major parent ^c	Type of recombination	Haplotype	Analysis in:
	Begin	End						
1	1835	33	MSV-A_ZA_Omr_g221-2007 ^b	MSV-A_ZA_Por1_D14-2008 ^b	MSV-A_ZW_He12_Bet36-2006 ^b	Intra-MSV-A ₁	MSV-A ₁ I	This study; 70
2	144	1674	MSV-A_ZA_Hec4_O13-1989	MSV-A_UG_Bug245-2005 ^b	MSV-A_MZ_Bill_Moz20-2007 ^b	Intra-MSV-A ₁	MSV-A ₁ XII	This study
3	2667	1595	MSV-A_UG_Jin219-2005 ^b	MSV-A_UG_Luw196-2005 ^b	Unsampled	Intersubtype (MSV-A ₂ -A ₄)	MSV-A ₁ VIII	This study; 47, 70
4	1961	78	MSV-A_ZA_TheE_g132-2006 ^b	MSV-A_ZA_Tra_D4-2006 ^b	MSV-A_NG_Iba3_N16-2006 ^b	Intra-MSV-A ₁	MSV-A ₁ XI	This study; 70
5	1	1573	MSV-A_NG_Oko_N32a-2007 ^b	MSV-A_UG_Mba1308-2005 ^b	Unsampled	Intersubtype (MSV-A ₂ -A ₄)	MSV-A ₁ VIII	This study; 47, 70
6	2241	641	MSV-A_ZW_MatC-1989	MSV-A_UG_Nak120-2005 ^b	Unsampled (MSV-A ₁) ^b	Intra-MSV-A ₁	MSV-A ₁ XIII	This study; 47, 70
7	1741	2684	MSV-A_ZA_Let5_M46K-1989 ^b	MSV-A_MZ_Map4_Moz25-2007 ^b	MSV-A_UG_Bush53-2005 ^b	Intra-MSV-A ₁	MSV-A ₁ XIV	This study
8	1993	2658	MSV-A_CF_Bang4_Car42-2008 ^b	Unsampled	MSV-A_UG_Bug245-2005 ^b	Distant from MSV-A ₁	MSV-A ₁ IV	This study; 47
9	1272	2210	MSV-A_ZW_Maz4_Mic3-1992	Unsampled (MSV-A_ZW_Mas5_Mic7-1993)	MSV-A_ZW_Mas2_Mic4-1993 ^b	Distant from MSV-A ₂ -A ₃ -A ₄	MSV-A ₁ X	This study
10	1318	33	MSV-A_UG_Mub94-2005 ^b	Unsampled (MSV-A_ZW_Mas5_Mic7-1993)	MSV-A_ZM_Ch3_Z8-2008 ^b	Distant from MSV-A ₁ -A ₂ -A ₃ -A ₄	MSV-A ₁ I	This study; 47
11	2646	1301	MSV-A_KE_Nye2_Ken12-2008	Unsampled (MSV-A_UG_Iga231-2005 ^b)	Unsampled	Distant from MSV-A ₁ -A ₂ -A ₃ -A ₄	MSV-A ₁ IX	This study
12	1972	2375	MSV-A_KE_Nan3_Ke8-2008 ^b	MSV-A_ZA_Let5_M46K-1989	MSV-A_NG_Odo_N18-2006 ^b	Distant from MSV-A ₂ -A ₃ -A ₄	MSV-A ₁ V	This study
13	1580	36	MSV-A_BE_Oua1_BF3-2008 ^b	Unsampled (MSV-A_ZW_Mas5_Mic7-1993)	MSV-A_CF_Bang10_Car49-2008 ^b	Distant from MSV-A ₂ -A ₃ -A ₄	MSV-A ₁ VI	This study; 47, 70
14	35	1239	MSV-A_ZA_Pot9_O34-1979 ^b	MSV-A_ZA_Na12_g194-2007 ^b	MSV-A_ZA_Por2_O26-1979 ^b	Distant from MSV-A ₂ -A ₃ -A ₄	MSV-A ₁ III	This study
16	551	1833	MSV-A_ZA_He1_O9-1979 ^b	MSV-A_ZW_He12_Bet36-2006	Unsampled (MSV-A_ZA_Por2_O26-1979 ^b)	Distant from MSV-A ₁ -A ₂ -A ₃ -A ₄	MSV-A ₁ II	This study
17	2162	130	MSV-A_ZW_Mas5_Mic7-1993	MSV-A_ZW_Mas2_Mic4-1993 ^b	MSV-A_ZA_Geo_M51-1988	Intersubtype (MSV-A ₂ -A ₃ -A ₄)	MSV-A ₁ XVI	This study
2 ⁸	2337	2626	MSV-A_ZA_MakD 1989 ^b	Unsampled	MSV-A_ZW_MarB-1996 ^b	Distant from MSV-A ₂ -A ₃ -A ₄	MSV-A ₁ II	47
3 ⁸	1616	2626	MSV-A_UG_Kas70-2005 ^b	MSV-A_ZA_MakD 1989 ^b	MSV-A_UG_Luv107-2005 ^b	Distant from MSV-A ₂ -A ₃ -A ₄	MSV-A ₁ III	47
9 ⁸	1129	1945	MSV-A_KE_Ker ^b	MSV-A_KE_Km	MSV-A_KE_MiKA-1997	Intersubtype (MSV-A ₁ -A ₃)	MSV-A ₁ I	47

^a The nt positions (counted from the sequence TAATATTAC in the alignment [see data set S1]) reflect the boundaries of the strongest recombination signal detected. For those positions with an apostrophe, actual breakpoint sites may not correspond to the breakpoint position.

^b Multiple recombinants were inferred to share evidence of the same recombination event, or various other parental sequences were inferred to be equally related to the actual parental sequences.

^c Where the parental sequence is referred to as "unsampled," the MSV-A isolate within parentheses is identified as the potential minor or major parent or was used to infer the existence of an unsampled parental sequence.

FIG. 1. ML phylogeny of 353 fully sequenced MSV-A genomes. The MSV-A genomes were sampled from across Africa and on La Réunion Island. The phylogeny (model GTR+I+G₄) was rooted using the isolate MSV-B (ZA_PlaB_g27-2006) (data not shown). Except where indicated with an asterisk, branches occurring in <50% of 100 full ML bootstrap replicates were collapsed. MSV-A variant designations are consistent with those outlined previously (39, 47). Different colors or shades are used to denote noteworthy clades within the variant groupings. The linearized genome schematics illustrate the recombination mosaics used to further classify the isolates into recombinant-lineage groupings. The coloring is consistent with that used to describe variant groups and therefore denotes as accurately as possible the contributors of the sequence tracts of which the recombination mosaics are composed. The gray regions indicate tracts derived either from currently unsampled MSV-A variants or from ancestral sequences that were not obviously classifiable as belonging to the MSV-A₂, -A₃, or -A₄ lineage. The hatched regions indicate tracts of

degrees of movement observed for fairly common lineages, such as MSV-A₄V (zones 13 and 14 in Fig. 2M), corroborating the commonly observed localization of MSV-A₄ genomes to southern Africa (39, 69). Within southern Africa, the MSV-A₄V distribution differs from that of the other main MSV-A₄ recombinant lineage, MSV-A₄I. Besides having a more northeasterly localization (zone 11 in Fig. 2L), MSV-A₄I also differs from MSV-A₄V in its possession of a recombinationally derived tract of MSV-A₁-like sequence (Fig. 1). It is remarkable that MSV-A₄I has remained restricted to southern Africa while the apparently far younger MSV-A₁II and MSV-A₁V lineages (Fig. 2C and F, respectively) found alongside MSV-A₄I in southern Africa are far more widely distributed across the continent.

Determining where MSV-A first emerged. In an attempt to determine how the MSV-A subtype and recombinant-lineage distributions arose, we analyzed our FG and MPCP data sets using a suite of newly developed Bayesian phylogeographic techniques (35, 36) implemented in the program BEAST (11). These techniques use information on the locations where viruses were sampled to reconstruct, within a phylogenetic context, the pathways of past virus movements that gave rise to the observed sampling locations. Importantly, these analyses also use sampling dates and apply sophisticated molecular clock models to enable both the timing of virus movements and the inference of when and where any particular ancestral sequences might have arisen.

The time-calibrated phylogeographic history of MSV-A, as indicated by the FG and MPCP data sets, is summarized in Fig. 3 in the form of maximum clade credibility (MCC) trees and as animations that can be viewed using Google Earth (summarized in Fig. 4, the FG data set [see data set S2.kml in the supplemental material], and the MPCP data set [see data set S3.kml in the supplemental material]). These trees and animations were produced using the discrete phylogeographic model, and the trees have branches that are color coded to indicate the most probable locations, among those considered, of the ancestral viruses represented within the trees. Color transitions along branches indicate instances where ancestral sequences probably moved between countries.

Although mostly congruent, the two trees display one potentially important topological difference: whereas the MSV-A₆ isolates form a separate clade basal to the other MSV-A subtypes in the FG tree and which apparently diverged from the other MSV-A subtypes in approximately 1864 (95% highest probability density interval [HPD], 1809 to 1935), in the MPCP tree, the MSV-A₆ isolates occur nested within a clade including MSV-A₂, -A₃, and -A₄ isolates and are estimated to have diverged from these in approximately 1952 (Fig. 3). Importantly, when using the far less parameter-rich continuous-diffusion model, the MCC trees were topologically almost identical

to their counterparts produced using the discrete model (Fig. 3; see Fig. S2 in the supplemental material). While the differences between the trees seriously diminishes our confidence in the estimated timings of the split between MSV-A₆ and the other MSV subtypes, they are an inevitable consequence of the MPCP data set containing substantially less phylogenetic information than the FG data set. Importantly, this uncertainty was accounted for in the subsequent phylogeographic analyses, given the poorly supported branches within the MPCP tree (<0.4 posterior state probability [PSP]) for both of the tree branches between the MSV-A₆ split and the root node of the tree.

Despite these differences between the MCC trees constructed from the FG and MPCP data sets, the results of the discrete model analyses on both data sets indicated that, among the 14 countries from which we sampled, southern Africa in general (PSP for the countries Zambia [ZM], Zimbabwe [ZW], Mozambique [MZ], Lesotho [LS], and South Africa [ZA], 0.5287 for the FG data set and 0.6497 for the MPCP data set [Fig. 3]), and either Zimbabwe (FG data set, PSP = 0.298) or South Africa (MPCP data set, PSP = 0.319) in particular, was the most probable location of the MRCA of all the analyzed MSV-A isolates. A southern African origin for MSV-A was also inferred under the continuous models, with the entire 80% HPD posterior mass (representing the uncertainty on the inferred MCC tree root location) falling within the southern African region. This result was obtained irrespective of the data set investigated or the specific diffusion model employed (see Fig. S2 in the supplemental material).

Despite the fact that the spatial estimates for the MSV-A origin were consistent, the discrete and continuous model runs inferred slightly different dates for the MSV-A MRCA, with the former model yielding estimates of 1933 (95% HPD, 1867 to 1950) and 1863 (95% HPD, 1809 to 1935) for the MPCP and FG data sets, respectively. Similarly, the continuous model runs yielded MRCA date estimates of 1901 (95% HPD, 1828 to 1941) for the MPCP data set and 1887 (95% HPD, 1842 to 1939) for the FG data set (Fig. 3; see Fig. S2 in the supplemental material). Importantly, these estimated dates have broadly overlapping 95% HPDs encompassing the first credible reports of MSV-A-like disease in South Africa during the 1870s (17). These estimates are very similar to the MSV-A MRCA dates determined previously with much smaller data sets using far less parameter-rich models (21).

Identifying the major MSV-A migration pathways. The present geographical distributions of the various MSV-A recombinant lineages (Fig. 2) imply that there have potentially been quite complex patterns of MSV-A movement throughout the continent since the probable emergence of this strain in southern Africa during the mid-1800s.

Collectively, using the discrete phylogeographic model, the

sequence whose origins are possibly derived from either unsampled MSV-A₂- or -A₄-like lineages or ancestral viruses from before the MSV-A₂/A₄ split. The numbers on the recombination mosaics correspond to the recombination event numbers in the RDP file (see Table S1 in the supplemental material). Events marked with the symbol § are those identified by Owor et al. (47) only. Each mosaic pattern is assigned a recombinant-lineage name based (with slight modification) on the convention of Owor et al. (47). The diagram above the uppermost lineage genome schematic shows the positions of genomic features: V2, movement protein gene; V1, coat protein gene; SIR, short intergenic region; C1/C2, replication-associated protein gene; C1, *repA* gene. The bar indicates the number of nucleotide substitutions per site.

FIG. 2. The sampling locations, distributions, and current population prevalences of various MSV-A recombinant lineages. Twelve MSV-A lineages were sampled throughout Africa between the years 2000 and 2008. (A) The main maize-growing areas were split into 14 zones, and the total number of MSV-A isolates analyzed from each zone is noted. Zone 1 is West Africa, zone 2 is Central Africa, zones 3 to 10 are East Africa, and zones 11 to 14 are southern Africa. The blue circles indicate sampling sites, and the intensity of the color corresponds to the extent of sampling in that region. Multiple samples from any region, irrespective of their lineage designation, were considered in this analysis. (B to M) Linearized genomic schematics indicate the recombination mosaics of the various MSV-A recombinant lineages. The color scale illustrating population prevalence is exponential (base 2) from 3.125 to 50% and linear from 50 to 100% (47).

FIG. 3. MSV-A maximum clade credibility phylogenies generated under the discrete phylogeographic diffusion model. The MSV-A FG (A) and MPCP (B) phylogenies are color coded so that the different branch colors depict the most probable location states of their descendant nodes. These location states have been defined as geographical regions encompassing the following countries: West Africa (Benin [BJ], Cameroon [CM], Burkina Faso [BF], and Nigeria [NG]) in blue, Central Africa (Central African Republic [CF] and ZM) in gray, East Africa (Kenya [KE], Uganda [UG], and northern MZ) in green, and southern Africa (ZA, ZW, LS, and southern MZ) in red. The posterior probability distributions for the root state locations (i.e., indicating the most probable site of origin of the most recent common ancestor of all the sequences represented) are shown in the bottom left corner of each phylogeny, using the colors above the two-letter country codes (TD, Chad; RE, La Réunion). The vertical lines indicate the distribution of subtypes MSV-A₁, -A₂, -A₃, -A₄, and -A₆. YBP, years before present.

FG and MPCP data set analyses identified a total of 32 well-supported (i.e., with a BF of >3.2) MSV-A movement pathways, 8 of which were concordant between the FG and MPCP data sets (Fig. 4; see Table S1 in the supplemental material). Although approximately the same intercountry pathways were also inferred with both data sets when analyzed using the continuous-diffusion models, there were numerous additional intracountry movement pathways identified by the analyses employing this model (FG data set and MPCP data set [see data sets S4.kml and S5.kml in the supplemental material]).

All of the phylogeographic models indicated that, following the initial emergence of MSV-A in southern Africa, long-range northerly transcontinental MSV-A movements out of southern Africa in general, but most probably out of either Zimbabwe (BF = 19.6 for the FG data set and 5.9 for MPCP data set) or South Africa (BF = 2,046 for the FG data set and 4,338 for the MPCP data set), occurred throughout the 20th century (see data sets S2.kml, S3.kml, S4.kml, and S5.kml in the supplemental material, which can be viewed as animations in Google Earth). Based on our analysis of the FG data set with the discrete phylogeographic model, it is likely that the progenitor of the MSV-A₂ lineage reached West Africa between 1930 and 1968 (95% HPD, 1909 to 1979), the MSV-A₃ progenitor reached East Africa between 1915 and 1969 (95% HPD, 1888

to 1979), and the MSV-A₆ progenitor reached La Réunion Island between 1880 and 1970 (95% HPD, 1809 to 1981).

Due to the fact that our analysis considered only MSV isolates sampled since 1979 and to the inclusion in our analysis of only 12 isolates from La Réunion Island and none from other Indian Ocean islands, we were unable to unequivocally identify a plausible migratory route for the seeding of the island from mainland Africa. Both the FG and MPCP analyses weakly supported Zimbabwe (BF < 3) as being the source of the progenitor MSV-A₆ isolate that ultimately evolved into those sampled on La Réunion since the mid-1980s. The support for these movements is, however, extremely tenuous and provides us with no real insights into the actual geographical origins of the MSV-A₆ lineage, which possibly has been present on La Réunion since the 1920s (27, 28, 29).

In addition to the earliest transcontinental dissemination events during the initial diversification of MSV-A into what we today recognize as distinct subtypes, general patterns evident from Fig. 4 suggest recent and numerous localized movements of MSV-A₁ and -A₄ isolates within regions in which these subtypes are already established. There are, for example, well-supported MSV-A₁ (BF = 958 for the FG data set and BF = 419 for the MPCP data set) and -A₄ (BF = 108 for the FG data set and BF = 851 for the MPCP data set) movements among

FIG. 4. MSV-A migration routes inferred under the discrete phylogeographic diffusion model. MSV migration events were inferred using BEAST analyses and the FG and MPCP data sets. The countries from which samples were obtained are labeled with two-letter codes. Migration events are indicated using arrows. The arrow colors depict the mean ages (in years) of the individual movement events that they represent. Whereas movement events inferred with both the FG and MPCP data sets are represented using solid lines, events inferred with only the MPCP or FG data set are represented with dotted and dashed lines, respectively. The thicknesses of the lines indicate the BF values associated with particular movement pathways (with each BF representing all individual movements between two locations). In cases where both the MPCP and FG data sets yielded BFs of >3.2 for a particular movement pathway between any two locations, the thickness of the line indicates the higher of the two BFs.

southern African countries, northerly movements of MSV-A₁ viruses from southern Africa to east Africa (BF = 15.1 for the FG data set), MSV-A₁ movements from east Africa to southern Africa (BF = 4.81 for the MPCP data set and 80,036 for the FG data set) (see Table S1 in the supplemental material), and MSV-A₁ movements westward from east Africa to central Africa (BF = 4.46 for the FG data set and BF = 7,679 for the MPCP data set).

It is notable that besides an extremely well supported movement from Nigeria to central Africa (BF = 6,049 and BF = 80,036 for the MPCP and FG data sets, respectively), these analyses revealed no evidence of MSV-A movements from West Africa to either East or southern Africa using either the FG or MPCP data set.

While similar patterns of long- and short-range movements were inferred with both the FG and MPCP data sets, the timings of specific movements inferred with the MPCP data set tended to be more recent than those inferred with the FG data set (see Fig. S2 in the supplemental material). Besides superficial differences in the timings of ancestral sequences and movement events, however, the inferred spatiotemporal dynamics of MSV-A dispersal during the 20th century were broadly concordant (see data sets S2.kml, S3.kml, S4.kml, and S5.kml in the supplemental material) across both data sets analyzed with both the discrete and continuous phylogeographic models. For example, both the FG and MPCP analyses indicated that southeast Africa is the region where MSV-A is likely to have initially emerged and the main launch pad of detectable transcontinental movements throughout the history

of MSV-A. The analyses also implicated East Africa in general, and Uganda in particular, as a secondary source of more recent MSV-A₁ movements throughout the continent.

Determining where and when the different MSV-A recombinant lineages arose. While our analysis indicates that a set of transcontinental movements, probably at some time during the early to mid-1900s, served to establish the spatial distributions of the various MSV-A subtypes found across Africa today, determining when and where the MRCAs of the currently circulating MSV-A recombinant lineages arose is probably of more immediate epidemiological interest. Of the 12 currently circulating MSV-A recombinant lineages (i.e., those that have been sampled between 2000 and 2008), the continuous phylogeographic model analyses of both the FG and MPCP data sets indicated that 6 had arisen in East Africa between 1975 and 1998 (with MSV-A₁VII being the oldest and MSV-A₁VIII the youngest) and a further 6 in the region around the intersection between South Africa, Zimbabwe, and Mozambique between 1947 and 1996 (with MSV-A₄V being the oldest and MSV-A₁XI the youngest) (Fig. 5). Importantly, the inferred locations of the various MRCAs determined using the FG and MPCP data sets (the 80% HPD posterior masses for the inferred locations of these ancestral viruses are indicated, respectively, by green and red lines in Fig. 5) were nearly identical, and both concurred that, relative to the other parts of Africa from which we have sampled viruses, East Africa and the eastern parts of southern Africa have played a disproportionately large role in the evolution and emergence of epidemiologically relevant MSV-A lineages.

FIG. 5. Geographical and temporal origins of various MSV-A recombinant lineages inferred using the continuous phylogeographic model. The origins of the 12 MSV-A lineages with members that were sampled between 2000 and 2008 (i.e., those represented in Fig. 2) were inferred using the FG and MPCP data sets. (A) The sampling densities and geographical distributions of viruses characterized as belonging to the various recombinant lineages are shown using blue circles. (B to M) Recombinant-lineage names, t_{MRCAs} of all sampled members of the lineages, and associated 95% HPDs of these estimated dates (as determined using the FG data set). The green and red lines on the maps delineate the 80% HPDs representing the uncertainty of the locations of the MRCAs of the various recombinant lineage groupings as inferred using the FG and MPCP data sets, respectively. The blue arrows on some of the maps indicate the apparent movements of descendants of the various recombinant-lineage MRCAs toward regions where these lineages are currently found.

Comparing the maps that show the estimated times and places where these MSV-A recombinant lineages probably arose (Fig. 5) with those that show the current distributions and population prevalences of these lineages (Fig. 2) also pro-

vides some insight into the movement dynamics of these important MSV-A lineages. For example, whereas three out of four of the MSV-A₁ recombinant lineages that probably arose in southern Africa (MSV-A₁II, -VI, and -XV) display clear

evidence of transcontinental northward movement during the past 50 years, neither of the MSV-A₄ recombinant lineages that emerged in the same region over an equivalent time period (MSV-A₄I and -V) display any evidence of similar long-distance movements (Fig. 2L and M).

An interesting example of the epidemiological dynamics of MSV-A that have probably occurred across East and West Africa during the last 30 years is given by MSV-A₁III and MSV-A₁VII (Fig. 2D and H and Fig. 5D and H). Despite appearances to the contrary given their different current distributions, both MSV-A₁VII and MSV-A₁III apparently emerged in East Africa and moved toward West Africa. However, whereas MSV-A₁VII emerged in East Africa around 1975 (95% HPD, 1967 to 1983), it is presently the dominant MSV-A recombinant lineage in West Africa and is totally undetectable in East Africa. MSV-A₁III, on the other hand, is inferred to have emerged in East Africa only around 1994 (95% HPD, 1988 to 1998) and has become the dominant MSV-A lineage in this region while it remains only barely detectable in West Africa. This suggests that over decades there might have been waves of east to west virus movements where epidemiologically important viral lineages continually emerge in East Africa and move westward while concomitantly being replaced in East Africa by newer lineages.

Given the estimated timing of the most recent common ancestors (t_{MRCA}) of the various MSV-A recombinant lineages, it is evident that over the past 60 years East Africa and the eastern parts of southern Africa have seen the repeated genesis of novel epidemiologically important lineages. Although many of these have seemingly moved quite long distances from their sites of origin, the continuing presence of older recombinant lineages, such as MSV-A₁II and MSV-A₄V, at their sites of origin indicates that, in spite of the transient “boom and bust” persistence profiles of lineages such as MSV-A₁VII, lineages such as MSV-A₁II and MSV-A₄V may have different epidemiological characteristics that have provided them with a greater degree of persistence within the source population at their geographical origin.

Notwithstanding the apparent differences in migratory potential displayed by different MSV-A recombinant lineages, the two general patterns indicated by our analyses are, first, that there are likely few physical barriers to the spread of MSV-A throughout the major maize-growing regions of continental Africa and, second, that it can take as little as 30 years for a novel MSV-A lineage to traverse the continent.

Crucially, the analyses of the FG and MPCP data sets with the continuous phylogeographic model also furnished us with estimates of the mean rates at which different MSV-A lineages have been moving over the past century (see Fig. S3 in the supplemental material). Whereas the branch lengths in these trees represent elapsed time, the branch colors represent relative movement rates. The lowest mean rates inferred using the FG data set (22.8 km/year; 95% HPD, 4.5 to 45.5) were estimated along the most basal branches in the phylogeny (i.e., that occurring during the initial spread in southwest of Africa in the first few decades of the 20th century). Intermediate rates (34.0 km/year; 95% HPD, 17.3 to 53.4) were inferred over the rest of the 20th century, during which the southern African founder population probably spread to the rest of Africa. The rates of short-range MSV-A movement over the past decade

have been ~48.8 km/year (95% HPD, 38.7 to 59.7). While this result may seem to indicate that MSV-A movement rates are increasing, it is important to note that the HPDs of all three estimates are very broadly overlapping and are probably not significantly different from one another. In fact, despite the different root date estimates made for the MPCP and FG MCC trees, the estimated mean MSV-A movement rates are almost identical across both of these trees (i.e., over the entire evolutionary history of MSV-A as expressed using these trees): for the MPCP data set, it is 32.4 km/year (95% HPD, 14.90 to 51.34), and for the FG data set, it is 32.6 km/year (95% HPD, 16.28 to 51.90).

Concluding remarks. Although we were able to infer both the most probable origins of MSV-A and its transcontinental movement pathways, some limitations are inherent in our analyses and the conclusions that can be drawn from them. For example, our analyses by necessity considered only sites from which viruses have been sampled. It remains possible, however, that vast unsampled areas of Africa where maize is not intensively cultivated (such as the Democratic Republic of Congo or Angola) could have been either the actual sites where various MSV-A ancestral sequences arose or the actual departure points of many of the historical migratory pathways indicated here. It is nevertheless noteworthy that our analyses using the continuous phylogeographic model identified southeastern Africa as the most likely origin of both the MSV-A MRCA and the MRCAs of 6 of the 12 main MSV-A recombinant lineages known to be currently circulating throughout the continent. Therefore, even if this region is not the ultimate origin of these lineages, it has almost certainly been a major gateway during their transcontinental movements.

From an epidemiological perspective, it would be of great interest to determine the factors underlying both the local and transcontinental MSV-A movement pathways that we have inferred. Although the movements most probably reflect the movement dynamics of the *Cicadulina* species and biotypes that are the most successful MSV-A vectors (1, 45), differences between the movement dynamics of MSV-A (and particularly MSV-A₁) and those of other MSVs and other African mastrevirus species in general (all of which are transmitted by these same leafhopper species [70]) provide support for the hypothesis that there exists a specific virus-encoded determinant of increased MSV-A mobility (69, 70). In this regard, it remains to be determined how meteorological factors, *Cicadulina* sp. host preferences (for either feeding or reproduction), the distributions and population densities of both virus and *Cicadulina* host species, and regional variations in both maize cultivation patterns and the use of MSV-resistant maize genotypes might have combined to yield the MSV-A movement pathways inferred by our analyses.

ACKNOWLEDGMENTS

We acknowledge Peter van Heusden (South African National Bioinformatics Institute) for his help and expertise in setting up the computational analyses required to run BEAST.

This project was partially funded by the National Research Foundation of South Africa (NRF-SA). A.L.M. was funded by the University of Cape Town, the NRF-SA, and the Canon Collins Trust. A.V. was funded by the Carnegie Corporation of New York. D.P.M. was funded by the Wellcome Trust. P. Lemey was funded by a postdoctoral fellowship from the Fund for Scientific Research (FWO) Flanders. P.

Lefevre was funded by the GIS Centre de Recherche et de Veille Sanitaire sur les Maladies Emergentes dans l'Océan Indien (CRVOI) through the EmerGe grant (no. PRAO/AIRD/CRVOI/08/03). J.-M.L. was funded by the Conseil Régional de la Réunion.

A.L.M., G.W.H., A.V., P. Lefevre, and D.P.M. conceived and designed the project, analyzed and interpreted the data, and drafted the manuscript. P. Lemey analyzed and interpreted the data and edited the manuscript. A.L.M., D.P.M., A.V., E.P.R., R.W.B., D.N.S., B.E.O., J.B.E., R.S., S.O., M.S., I.Z., E.K.K., D.P.L., N.M., J.M., S.S., A.T., F.T., T.V.A., J.-M.L., P.G.M., B.F., M.R., and O.P.W. acquired, cloned, and sequenced virus genomes.

REFERENCES

- Asanzi, M. C., N. A. Bosque-Pérez, L. R. Nault, D. T. Gordon, and G. Thottappilly. 1995. Biology of *Cicadulina* spp. (Homoptera: Cicadellidae) and transmission of maize streak virus. *Afr. Entomol.* **3**:173–179.
- Asanzi, C. M., N. A. Bosque-Pérez, and L. R. Nault. 1995. Movement of *Cicadulina storeyi* (Homoptera: Cicadellidae) in maize fields and its behaviour in relation to maize growth stage. *Ins. Sci. Appl.* **16**:39–44.
- Bigirwa, G., et al. 1995. A new maize disorder in Uganda caused by *Cicadulina niger*, p. 202–204. In D. C. Jewell, S. R. Waddington, J. K. Ransom, and K. V. Pixley (ed.), *Maize research for stress environments*. Proceedings of the Fourth Eastern and Southern Africa Regional Maize Conference, Harare, Zimbabwe. CIMMYT, Mexico City, Mexico.
- Boni, M. F., D. Posada, and M. W. Feldman. 2007. An exact non-parametric method for inferring mosaic structures in sequence triplets. *Genetics* **176**:1035–1047.
- Bosque-Pérez, N. A. 2000. Eight decades of maize streak virus research. *Virus Res.* **71**:107–121.
- Boulton, M. L., W. G. Buchholz, M. S. Marks, P. G. Markham, and J. W. Davies. 1989. Specificity of Agrobacterium-mediated delivery of maize streak virus DNA to members of the Gramineae. *Plant Mol. Biol.* **12**:31–40.
- Briddon, R. W., P. Lunness, L. C. Chamberlin, and P. G. Markham. 1994. Analysis of the genetic variability of maize streak virus. *Virus Genes* **9**:93–100.
- Bruen, T. C., H. Philippe, and D. Bryant. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**:2665–2681.
- Collin, S., M. Fernández-Lobato, P. S. Gooding, P. M. Mullineaux, and C. Fenoll. 1996. The two nonstructural proteins from wheat dwarf virus involved in viral gene expression and replication are retinoblastoma-binding proteins. *Virology* **219**:324–329.
- Dabrowski, Z. T. 1987. Two new species of *Cicadulina* China (Hemiptera: Euscelidae) from West Africa. *Bull. Entomol. Res.* **77**:53–56.
- Drummond, A., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**:214.
- Erdmann, J. B., et al. 2010. Replicative intermediates of Maize streak virus found during leaf development. *J. Gen. Virol.* **91**:1077–1081.
- Fauquet, C. M., et al. 2008. Geminivirus strain demarcation and nomenclature. *Arch. Virol.* **153**:783–821.
- Fennah, R. G. 1959. A new species of *Cicadulina* (Homoptera: Cicadellidae) from East Africa. *Ann. Mag. Nat. Hist. Ser.* **13**:757–758.
- Food and Agriculture Organization of the United Nations. 2007. reference year. Food balance sheets. Food and Agriculture Organization of the United Nations, Rome, Italy. <http://faostat.fao.org/site/368/DesktopDefault.aspx?PageID=368#ancor>.
- Firth, C., M. A. Charleston, S. Duffy, B. Shapiro, and E. C. Holmes. 2009. Insights into the evolutionary history of an emerging livestock pathogen: Porcine Circovirus 2. *J. Virol.* **83**:12813–12821.
- Fuller, C. 1901. Mealie variegation. First entomology report of the Government of Natal (South Africa) 1899–1900, p. 17–19. P. Davis & Sons, Pietermaritzburg, South Africa.
- Gibbs, M. J., J. S. Armstrong, and A. J. Gibbs. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**:573–582.
- Grasso, C., and C. Lee. 2004. Combining partial order alignment and progressive multiple sequence alignment increases alignment speed and scalability to very large alignment problems. *Bioinformatics* **20**:1546–1556.
- Guindon, S., and O. Gascuel. 2003. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**:696–704.
- Harkins, G. W., et al. 2009. Dating the origins of the maize-adapted strain of maize streak virus, MSV-A. *J. Gen. Virol.* **90**:3066–3074.
- Hefferon, K. L., Y.-S. Moon, and Y. Fan. 2006. Multi-tasking of nonstructural gene products is required for bean yellow dwarf geminivirus transcriptional regulation. *FEBS J.* **273**:4482–4494.
- Hofer, J. M. I., et al. 1992. Coordinate regulation of replication and virion sense gene expression in wheat dwarf virus. *Plant Cell* **4**:213–223.
- Jeffreys, M. D. W. 1963. How ancient is West African maize? *J. Int. African Inst.* **33**:115–131.
- Jeske, H., M. Lutgemeier, and W. Preiss. 2001. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *EMBO J.* **20**:6158–6167.
- Kass, R., and A. Raftery. 1995. Bayes factors. *J. Am. Stat. Assoc.* **90**:773–795.
- Kopp, A. 1930. Les maladies des plantes à La Réunion. *Rev. Bot. Appl. Agric. Trop.* **105**:1–7. (In French.)
- Kopp, A., and D. d'Emmerez de Charmoy. 1931. Nouvelles constatations sur les maladies à virus de la canne à sucre et du maïs. *C. R. Acad. Sci.* **193**:876–878. (In French.)
- Kopp, A., and D. d'Emmerez de Charmoy. 1932. Observations nouvelles concernant la mosaïque de la canne à sucre et le streak du maïs à la Réunion. *Station Agronomique Réunion. Travaux Techniques* **3**:1–10. (In French.)
- Lam, T. T.-Y., et al. 2008. Evolutionary and transmission dynamics of reasortant H5N1 influenza virus in Indonesia. *PLoS Pathog.* **4**:e1000130.
- Lawry, R., D. P. Martin, D. N. Shepherd, T. van Antwerpen, and A. Varsani. 2009. A novel sugarcane-infecting mastrevirus from South Africa. *Arch. Virol.* **154**:1699–1703.
- Lazarowitz, S. G., and A. J. Pinder. 1989. Molecular genetics of Maize streak virus. *UCLA Symp. Mol. Cell. Biol.* **101**:167–183.
- Lazarowitz, S. G., A. J. Pinder, V. D. Damsteegt, and S. G. Rogers. 1989. Maize streak virus genes essential for systemic spread and symptom development. *EMBO J.* **8**:1023–1032.
- Lefevre, P., J.-M. Lett, B. Reynaud, and D. P. Martin. 2007. Avoidance of protein fold disruption in natural virus recombinants. *PLoS Pathog.* **3**:e181.
- Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard. 2009. Bayesian phylogeography finds its roots. *PLoS Comp. Biol.* **5**:e1000520.
- Lemey, P., A. Rambaut, J. J. Welch, and M. A. Suchard. 2010. Phylogeography takes a relaxed random walk in continuous space and time. *Mol. Biol. Evol.* **27**:1877–1885.
- Markov, P. V., et al. 2009. Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. *J. Gen. Virol.* **90**:2086–2096.
- Martin, D. P., and E. P. Rybicki. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**:562–563.
- Martin, D. P., et al. 2001. Sequence diversity and virulence in Zea mays of Maize streak virus isolates. *Virology* **288**:247–255.
- Martin, D. P., and E. P. Rybicki. 2002. Investigation of maize streak virus pathogenicity determinants using chimaeric genomes. *Virology* **300**:180–188.
- Martin, D. P., D. Posada, K. A. Crandall, and C. Williamson. 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombinant breakpoints. *AIDS Res. Hum. Retrovir.* **21**:98–102.
- Martin, D. P., and D. N. Shepherd. 2009. The epidemiology, economic impact and control of maize streak disease. *Food Security* **1**:305–315.
- Martin, D. P., et al. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**:2462–2463.
- McCann, J. 2001. Maize and grace: history, corn, and Africa's new landscapes, 1500–1999. *Comp. Stud. Soc. Hist.* **43**:246–272.
- Okoth, V. A. O., and Z. T. Dabrowski. 1987. Population density, species composition and infectivity with maize streak virus (MSV) of *Cicadulina* spp. leafhoppers in some ecological zones in Nigeria. *Acta Oecol. Oecol. Appl.* **8**:191–200.
- Okoth, V. A. O., Z. T. Dabrowski, G. Thottappilly, and H. F. Van Emden. 1988. Comparative analysis of some parameters affecting maize streak virus (MSV) transmission of various *Cicadulina* spp. populations. *Inst. Sci. Appl.* **8**:445–461.
- Owor, B. E., et al. 2007. Genetic analysis of maize streak virus isolates from Uganda reveals widespread distribution of a recombinant variant. *J. Gen. Virol.* **88**:3154–3165.
- Owor, B. E., et al. 2007. Successful application of FTA Classic Card technology and use of bacteriophage phi29 DNA polymerase for large-scale field sampling and cloning of complete maize streak virus genomes. *J. Virol. Methods* **140**:100–105.
- Padidam, M., S. Sawyer, and C. M. Fauquet. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**:218–225.
- Peterschmitt, M., M. Granier, R. Frutos, and B. Reynaud. 1996. Infectivity and complete nucleotide sequence of the genome of a genetically distinct strain of maize streak virus from Réunion Island. *Arch. Virol.* **141**:1637–1650.
- Posada, D., and K. A. Crandall. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **98**:13757–13762.
- Prasanna, H. C., and M. Rai. 2007. Detection and frequency of recombination in tomato-infecting begomoviruses of South and Southeast Asia. *Virol. J.* **4**:111.
- Rambaut, A., and A. J. Drummond. 2007. Tracer v1.4, Available from <http://beast.bio.ed.ac.uk/Tracer>.
- Reynaud, B. 1988. Transmission des virus de la striure, du stripe et de la mosaïque du maïs par leurs vecteurs *Cicadulina mbila* (Naudé) et *Peregrinus maidis* (Ashmead). Approches biologique, génétique et épidémiologique de la relation vecteur-virus-plante. Ph.D. thesis. Université des Sciences et Techniques du Languedoc, Montpellier, France.
- Rose, D. J. W. 1962. Insect vectors of maize streak. *Zool. Soc. S. Afr. News Bull.* **3**:11.

56. **Rose, D. J. W.** 1972. Times and sizes of dispersal flight by *Cicadulina* spp. vectors of maize streak disease. *J. Anim. Ecol.* **41**:494–506.
57. **Rybicki, E. P., S. Dennis, G. Napier, and F. L. Hughes.** 1998. Novel grass and wheat strains of Maize streak virus detected by DNA amplification and sequencing. *Mol. Plant Pathol.* On-Line <http://www.bspp.org.uk/mppol/1998/0929rybicki/index.htm>.
58. **Saunders, K., A. Lucy, and J. Stanley.** 1992. RNA-primed complementary-sense DNA synthesis of the geminivirus African cassava mosaic virus. *Nucleic Acids Res.* **20**:6311–6315.
59. **Schnippenkoetter, W. H., et al.** 2001. The relative infectivities and genomic characterisation of three distinct mastreviruses from South Africa. *Arch. Virol.* **146**:1075–1088.
60. **Shepherd, D. N., et al.** 2008. A protocol for the rapid isolation of full geminivirus genomes from dried plant tissue. *J. Virol. Methods* **149**:97–102.
61. **Shepherd, D. N., et al.** 2010. Maize streak virus: an old and complex 'emerging' pathogen. *Mol. Plant Pathol.* **11**:1–12.
62. **Smith, J. M.** 1992. Analysing the mosaic structure of genes. *J. Mol. Evol.* **34**:126–129.
63. **Smith, G. J., et al.** 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* **459**:1122–1125.
64. **Storey, H. H.** 1925. The transmission of streak disease of maize by the leafhopper *Balclutha mbila* Naudé. *Ann. Appl. Biol.* **12**:422–439.
65. **Suchard, M. A., R. E. Weiss, and J. S. Sinsheimer.** 2001. Bayesian selection of continuous time Markov chain evolutionary models. *Mol. Biol. Evol.* **18**:1001–1013.
66. **Talbi, C., et al.** 2010. Phylogenetics and human-mediated dispersal of a zoonotic virus. *PLoS Pathog.* **6**:e1001166.
67. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. Mega4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**:1596–1599.
68. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
69. **Varsani, A., et al.** 2008. Recombination, decreased host specificity and increased mobility may have driven the emergence of maize streak virus as an agricultural pathogen. *J. Gen. Virol.* **89**:2063–2074.
70. **Varsani, A., et al.** 2009. Comparative analysis of Panicum streak virus and Maize streak virus diversity, recombination patterns and phylogeography. *Virol. J.* **6**:194.
71. **Willment, J. A., D. P. Martin, and E. P. Rybicki.** 2001. Analysis of the diversity of African streak mastreviruses using PCR-generated RFLPs and partial sequence data. *J. Virol. Methods* **93**:75–87.
72. **Woolston, C. J., H. V. Reynolds, N. J. Stacey, and P. M. Mullineaux.** 1989. Replication of Wheat dwarf virus DNA in protoplasts and analysis of coat protein mutants in protoplasts and plants. *Nucleic Acids Res.* **17**:6029–6041.
73. **Zhan, X., K. A. Richardson, A. Haley, and B. A. M. Morris.** 1993. The activity of the coat protein promoter of chloris striate mosaic virus is enhanced by its own and C1-C2 gene products. *Virology* **193**:498–502.