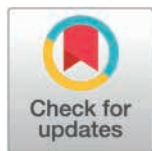RESEARCH ARTICLE

Grains, trade and war in the multimodal transmission of Rice yellow mottle virus: An historical and phylogeographical retrospective

Innocent Ndikumana[1◉], Geoffrey Onaga[2◉], Agnès Pinel-Galzi[3], Pauline Rocu[4], Judith Hubert[5,6], Hassan Karakacha Wéré[7], Antony Adego[7], Mariam Nyongesa Wéré[7], Nils Poulicard[3], Maxime Hebrard[8], Simon Dellicour[9,10,11], Philippe Lemey[10], Erik Gilbert[12], Marie-José Dugué[13], François Chevenet[14], Paul Bastide[15], Stéphane Guindon[4], Denis Fargette[3◉*], Eugénie Hébrard[3◉*]

1 Rwanda Agriculture Board, Kigali, Rwanda, **2** Africa Rice Center, Bouaké, Cote d'Ivoire, **3** PHIM Plant Health Institute, Univ Montpellier, IRD, CIRAD, INRAE, Institut Agro, Montpellier, France, **4** Department of Computer Science, Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier, CNRS and Université de Montpellier, Montpellier, France, **5** Department of Crop Science and Horticulture, Morogoro, Tanzania, **6** World Vegetable Center, Eastern and Southern Africa, Arusha, Tanzania, **7** Masinde Muliro University of Science and Technology, Kakamega, Kenya, **8** Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore, **9** Spatial Epidemiology Lab, Université Libre de Bruxelles, Brussels, Belgium, **10** Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium, **11** Interuniversity Institute of Bioinformatics in Brussels, Université Libre de Bruxelles, Vrije Universiteit Brussel, Brussels, Belgium, **12** Department of History, Arkansas State University State University, Jonesboro, Arkansas, United States of America, **13** Agronomy and Farming Systems, Saint Mathieu de Tréviers, France, **14** MIVEGEC, IRD, CNRS, Université de Montpellier, Montpellier, France, **15** IMAG, Université de Montpellier, CNRS, Montpellier, France

◉ These authors contributed equally to this work.
* denis.fargette@ird.fr (DF); eugenie.hebrard@ird.fr (EH)

Abstract

Rice yellow mottle virus (RYMV) is a major pathogen of rice in Africa. RYMV has a narrow host range limited to rice and a few related poaceae species. We explore the links between the spread of RYMV in East Africa and rice history since the second half of the 19th century. The phylogeography of RYMV in East Africa was reconstructed from coat protein gene sequences (ORF4) of 335 isolates sampled over two million square kilometers between 1966 and 2020. Dispersal patterns obtained from ORF2a and ORF2b, and full-length sequences converged to the same scenario. The following imprints of rice cultivation on RYMV epidemiology were unveiled. RYMV emerged in the middle of the 19th century in the Eastern Arc Mountains where slash-and-burn rice cultivation was practiced. Several spillovers from wild hosts to cultivated rice occurred. RYMV was then rapidly introduced into the nearby large rice growing Kilombero valley and Morogoro region. Harvested seeds are contaminated by debris of virus infected plants that subsist after threshing and winnowing. Long-distance dispersal of RYMV is consistent (i) with rice introduction along the caravan routes from the Indian Ocean Coast to Lake Victoria in the second half of the

19th century, (ii) seed movement from East Africa to West Africa at the end of the 19th century, from Lake Victoria to the north of Ethiopia in the second half of the 20th century and to Madagascar at the end of the 20th century, (iii) and, unexpectedly, with rice transport at the end of the First World War as a troop staple food from the Kilombero valley towards the South of Lake Malawi. Overall, RYMV dispersal was associated to a broad range of human activities, some unsuspected. Consequently, RYMV has a wide dispersal capacity. Its dispersal metrics estimated from phylogeographic reconstructions are similar to those of highly mobile zoonotic viruses.

Author summary

Rice yellow mottle virus (RYMV) poses a major threat to rice production in Africa. We explored through a multidisciplinary approach the links between the history of rice in East Africa since the second half of the 19th century and the spread of RYMV. The results illuminate the causes of RYMV diffusion. We show the role of long-distance caravan trade, the impact of the First World War and the consequences of seed exchange in the dispersal of RYMV. The paradoxical role of seeds in the spread of RYMV - which is vector transmitted and not seed transmitted – is explained in the light of rice biology and agronomy. Overall, this study reveals the wide range of transmission ways, some unexpected, in the dispersal of plant viruses. It also highlights the role of human transmission of pathogens and sheds light on the risk of transmission of RYMV and of other plant viruses from Africa to other continents.

Introduction

Contact transmission is a major transmission pathway for plant viruses, yet it has largely been neglected [1]. Contact-transmitted viruses have particles that are stable outside the infected cells, remain infectious on contaminated surfaces, and reach high concentrations (titer) in infected plants. Owing to their virion stability and high titer, Sobemoviruses (*Solemoviridae* family), Tobamoviruses (*Virgaviridae*), Potexviruses (*Alphaflexiviridae*) and Tombusviruses (*Tombusviridae*) are listed as candidates for contact and abiotic transmission [1]. The sobemovirus genus differs from the three other genera by having a narrow host range and by being transmitted by insect vectors. Beetles are the most frequently listed in sobemovirus transmission. Beetles are also involved in the transmission of several other genera [2]. This study aimed at evaluating the relative importance of abiotic and vector-mediated transmission in plant virus species with a dual mode of transmission (i.e., vector and abiotic). We hypothesized that the contribution of abiotic transmission to plant virus spread has been substantially underestimated [1,2]. Rice yellow mottle virus (RYMV), which affects rice crops in Africa, is an appropriate model to test this hypothesis.

Rice yellow mottle virus (RYMV) is a single-stranded RNA virus of the *Sobemovirus* genus [3]. RYMV is a major pathogen in all sub-Saharan African countries that grow rice [4]. Its host range is limited to the two cultivated rice species *Oryza sativa* and *O. glaberrima*, the wild rice species and a few related *Eragrostideae* species [5]. RYMV does not infect the seed embryo [6,7,8]. RYMV is transmitted by chewing insects, mainly beetles of the *Chrysomelidae* family in a non-persistent mode, and by leaf and root contact between infected and healthy plants [5]. RYMV reached high virus concentration within two to three weeks. It remains infectious in rice stubbles for months [5,7], and after passage through the intestinal tract of mammals [9]. The spread of the disease was first attributed to vector transmission, although no clear link has been found between the size of the beetle populations and the incidence of the disease. Later, the role of mechanical transmission through leaf-to-leaf and root-to-root contact was established experimentally [10]. Contamination in the soil also occurred by contact of the roots of young seedlings with infected rice stubble present in the soil. This is referred to as soil transmission or soil contamination [1,11,12]. Soil contamination has several origins, infected rice residues buried in the soil after harvesting or after mammal consumption, and also by planting rice seeds contaminated by infected plant residues (see below). Due to the stability of RYMV [7], soil transmission persists months after the infected rice residues have been buried. RYMV is disseminated by transplanting young infected seedlings from nurseries to fields sometimes located a few kilometers away [13]. Altogether, contact and soil transmission accounts for on-site survival of the virus and its local spread.

Yet, the long-distance transmission of RYMV has not been elucidated. Without experimental evidence, it has been tentatively ascribed to movements of beetles during tornadoes, transmission by more proficient flyer insects such as grasshoppers [5], or transmission by seeds in specific circumstances [14,15]. The last of the three hypotheses, which posited transmission with seeds, relied on the following observations. RYMV is not present in the rice embryo but is detected in the glumes, and in the outer envelopes of the grain [8]. Rice threshing and winnowing does not eliminate all the outer husks and all the leaf and stem residues [16]. Accordingly, infected plant debris remain in bags of rice seeds. We referred to them as contaminated seeds to make the distinction with infected seeds [4,17]. Rice seeds are transported between villages and across regions for food consumption and for agronomical purposes [18]. Rice planting with contaminated seeds as well as consumption of contaminated rice seeds by mammals (including humans) result in burying infected rice residues in the soil. Rice planted in a contaminated soil is infected through mechanical contact of the roots of the young seedlings with the infected residues. Based on these observations, it was hypothesized that RYMV was transported with bags of contaminated rice seeds to new areas, potentially far away from its sources [4,17]. Accordingly, seed transport would be the primary mode, possibly the exclusive mode, of long distance dissemination of RYMV.

Nevertheless, long-distance transmission of plant viruses, be it by vectors or with seeds, has been difficult to establish [19,20], and this is true for RYMV as well. None of the hypotheses put forward for long-distance transmission of RYMV has been experimentally validated. Phylogeography offers new perspectives through the reconstruction of the spatiotemporal pattern of spread [21,22]. It unveils the underlying processes that shape the dynamics of spread, including rare and past events, and lead to hypotheses on the modes of transmission and the impact of host biology and history. RYMV is a measurably evolving population [23], and the timescale of the phylogeographic reconstructions is not affected by the time dependent rate phenomenon [24]. The relaxed random walk (RRW) model in continuous phylogeography allows for variation in dispersal rates across branches of the phylogeny [22]. This model provides flexibility to accommodate the different means of transmission when reconstructing the phylogenetic dispersal history of the virus.

In this study, the phylogeography of RYMV in East Africa was reconstructed from a comprehensive heterochronous sequence dataset. Given that RYMV has a narrow host range limited to rice and a few related species, it was expected that its dissemination has been shaped by rice history. Subsequently, we paid special attention to the history of rice in East Africa since the beginning of the 19th century. RYMV originates in East Africa but unlike in Madagascar and in West Africa, rice cultivation until the last decades remained mainly confined to lake shores, river banks and wetlands [25, 26, 27]. Rice had been cultivated along the Indian Ocean Coast for centuries but rice cultivation was restricted to a few scattered coastal sites [26]. In the second half of the 19th century, large plantations of rice were reported in the

Morogoro region [28] and documented in the Kilombero valley, upstream of the mouth of the coastal river Rufiji, 300 km from the Indian Ocean [29]. By the end of the 19th century, rice supplies from the Rufiji flood plains had emerged as a sufficiently important source for that area to acquire the nickname "Little Calcutta" [30,31,32]. In the middle of the 19th century, rice was introduced around the great lakes of East Africa (Lake Victoria, Lake Tanganyika and Lake Malawi) [33]. Rice cultivation became generalized at the end of the 20th century. Assessing the virus movements between these far apart rice-producing regions helps to identify the main modes of dissemination.

Dispersal statistics have been increasingly used to characterize and quantify the spread of viruses [34,35,36]. We assessed the dispersal capacity of RYMV and compare it to a range of zoonotic viruses [35]. Altogether, this study suggests that the transport of contaminated rice seed was decisive in the long-distance dissemination of RYMV, not only within East Africa, but also in the introduction of the virus from East Africa to West Africa and to Madagascar. It points the most probable routes and modes of dissemination and unveils the links with rice history. It further shows the role of a broad spectrum of human activities in RYMV dispersal, some of them unsuspected. This explains why RYMV exhibits extensive dissemination capabilities, similar to those of highly mobile zoonotic viruses.

Results

Temporal signal

BETS is a Bayesian evaluation method of the temporal signal which involves comparing the fit to the data of two models: a model in which the data are accompanied by the actual (heterochronous) sampling times, and a model in which the samples are constrained to be contemporaneous (isochronous) [37]. BETS analyses indicate that the increase in sample size of heterochronous sequences lead to a stronger temporal signal in West Africa (from 261 to 335 sequences) and in East Africa (from 240 to 335 sequences) (Table 1). The temporal signal was more pronounced in the set of West African sequences (reflected by a higher Bayes factor). The temporal signal of the two datasets was also tested using cluster randomization tests of the sampling dates. There was no overlap of the credibility intervals of the substitution rate calculated with sampled dates and those with randomized dates. However, the differences obtained were more pronounced with the WA335 dataset. The square of the correlation coefficient of the regression sampling date vs. distance to the root obtained with TreeTime was 0.07 with the WA335 dataset vs. 0.01 with the EA335 dataset. Altogether, the temporal signal was more pronounced in the WA335 dataset than in the EA335 dataset. The stronger temporal signal of the West African dataset probably results from a relatively uniform distribution of the sampling dates over the last thirty years whereas that of East African isolates was concentrated over the last 15 years. Short sampling periods are prone to unpurged transitory deleterious polymorphism leading to higher and biased substitution rates [38]. Consistently, a recent article reported that substitution rates of plant viruses decreased with increasing sampling time span [39]. Viral isolates in contaminated

Table 1. Temporal signals in the datasets assessed by BETS.

Datasets	Log marginal likelihood*		Bayes factor
	Isochronous	dated	
West-Africa			
WA261	-10495	-10409	86
WA335	-13546	-13379	167
East-Africa			
EA240	-11294	-11246	48
EA335	-14634	-14539	95

*estimated under a strict molecular clock model.

https://doi.org/10.1371/journal.ppat.1013168.t001

seeds do not accumulate nucleotide substitutions over time. These periods of stasis, both frequent and prolonged during the initial stages of RYMV epidemiology in East Africa (see below), may introduce evolutionary rate variation among viral lineages. Altogether, the weak temporal signal of the EA335 dataset leads to a biased high substitution rate and correlatively to an inconsistent age of the most recent common ancestor (TMRCA) of the sampled sequences, similar to that in West Africa. It was obviously too recent as, given the nested position of the West African lineage within the East African lineages (see below), the TMRCA in East Africa should to be substantially older than that in West Africa.

As the temporal signal from the East African corpus was not sufficient alone to allow a valid reconstruction of the spatio-temporal dispersion of the RYMV in East Africa, priors derived from the WA335 dataset were applied. Accordingly, for the EA335 phylogeographic reconstructions, a prior was specified on the mean rate under the uncorrelated log-normal relaxed molecular clock (the "ucld.mean" parameter), a normal distribution with a mean of 1.24×10^{-3} substitutions per site per year and standard deviation of 1.16×10^{-4}. An informative prior was also specified on the standard deviation of the uncorrelated log-normal relaxed clock (the "ucld.stdev" parameter), a normal distribution with a mean of 1.09×10^{-3} substitution per site per year and standard deviation of 2.87×10^{-4}.

Phylogeny of RYMV in East-Africa

The phylogeny of RYMV in East Africa reconstructed from 335 coat protein gene sequences is made of three major lineages named, respectively, S4, S5 and S6 lineages (Fig 1). The phylogenetic group containing the S4 and S5 lineages
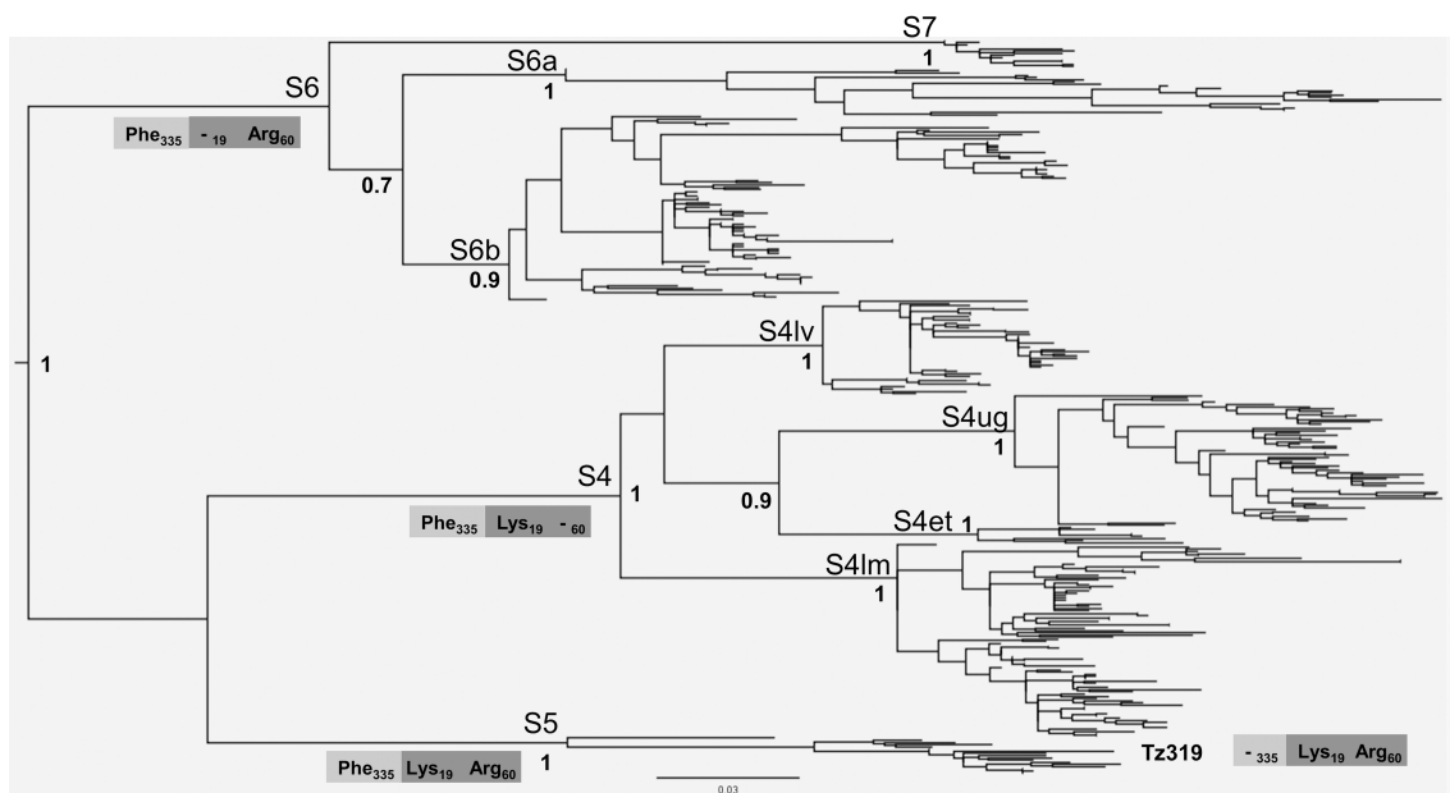
Fig 1. Phylogeny of RYMV in East Africa. Phylogeny of RYMV reconstructed from the coat protein gene sequences of 335 isolates through the maximum likelihood method under a HKY85 model and mid-point rooting. The names of the lineages and of the strains are given. The bootstrap support (> 0.70) is indicated at the node of the lineages and of the strains. The names of the lineages and of the strains are given. The insertion-deletion polymorphism (amino-acid and position) of the lineages and of the isolate Tz319 are framed in light grey for the ORF2a and in dark grey for the ORF4.

https://doi.org/10.1371/journal.ppat.1013168.g001

does not have a strong bootstrap support. Several strains are distinguished within the S4 and S6 lineages. The S4 lineage includes strains S4lm, S4lv, S4ug and S4et named according to their geographical distribution (lm, Lake Malawi; lv, Lake Victoria; ug, Uganda; et, Ethiopia). The S6 lineage includes strains S6a, S6b, and S7 but was not supported by a strong bootstrap value. The S5 lineage consists of one strain only.

Very few insertion-deletion events occurred in coding regions. The basal insertion-deletion pattern at positions 19 and 60 in the ORF4 in the S4, S5 and S6 lineages was that observed earlier [40]. An additional deletion occurred in the isolate Tz319 of the S5 strain at codon 432 in the ORF2a (Fig 1). It is the only insertion-deletion event found at a terminal phylogenetic position. All the insertion-deletion polymorphisms originated in isolates of the Eastern Arc Mountains.

Phylogeography from 335 coat protein gene sequences

Inference using statistical phylogeographic models, including those used in the present study, may be impacted by the way samples were collected [41,42]. In particular, observations deriving from a single point in time do not convey information about a directionality in the migration patterns, making it difficult, if not impossible, to accurately reconstruct ancestral locations. Biases in sampling locations, whereby these locations follow particular geographic patterns that depart from the distribution of the whole population, may also result in biases in ancestral reconstructions. The latter will reflect sampling locations rather than the spatial distribution of ancient populations. Finally, sampled locations are the product of two processes. The first governs the random fluctuation of lineage locations during the course of evolution. The second corresponds to logistic constraints that prevent specific regions from being included in the sampled areas. Ignoring these constraints leads to overly precise estimates of the spatial diffusion parameters [42]. The present study relies however on heterochronous data, which, at least in theory, conveys information about the directionality of the migration patterns. Also, it is likely that most regions of RYMV circulation in East Africa were surveyed and incorporated in our sample. Then, the sampling area should reflect the spatial distribution of the corresponding pathogen's population. The sampling scheme here is thus clearly different from targeting a small number of specific regions, selected based on logistic considerations only. While sampling biases cannot be fully excluded, their impact on the reconstructions of RYMV spread in Africa should be limited.

The spatio-temporal dispersal of RYMV lineages in East Africa was reconstructed from the capsid protein gene sequences of 335 isolates (Fig 2A and 2B), giving rise to the following most probable historical scenario.

(1) In the middle of the 19th century (1845, 95% HPD [1788–1897]), RYMV emerged around the Great Ruaha Escarpment in the Udzungwa mountains of the Eastern Arc Mountains north of the Kilombero Valley (Fig 2C);

(2) Within a few decades, RYMV dispersed to the adjacent Kilombero Valley (Fig 2C);

(3) Between the end of the 19th century and the beginning of the 20th century, RYMV spread to the Morogoro region 180 km at the north east of the Kilombero valley. During the same period, RYMV reached the south of Lake Victoria 400 km northward, the first long range dispersal event (I) (Fig 2D);

(4) In the following decades, RYMV circulated at the east of Lake Victoria (Fig 2E);

(5) A second long range dispersal event (II) originated at the beginning of the 20th century (1920, 95% HPD [1887–1949]) in the Kilombero valley or in the Morogoro region to reach the south of Lake Malawi, 900 km southward, ca. 80 years later (Fig 2E and 2F);

(6) A third long range dispersal event (III) took place in the second half of the 20th century. It originated at the east of Lake Victoria and reached northern Ethiopia, 1400 km northward, 50 years later (Fig 2F);
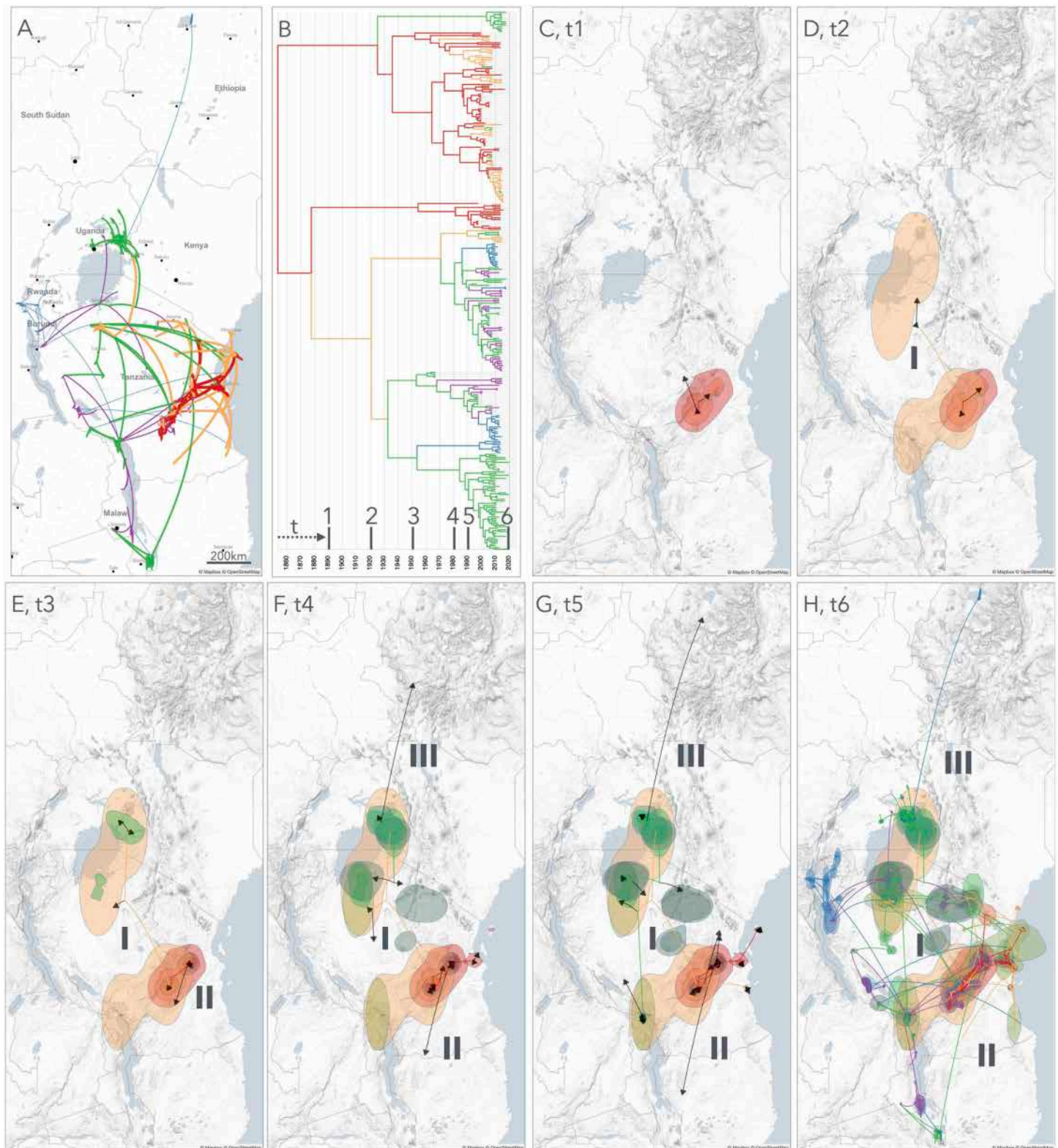
Fig 2. Phylogeography of RYMV in East Africa. Continuous phylogeographic reconstruction of the spatiotemporal dispersal of RYMV lineages in East Africa from 1850 to 2020 based on the coat protein gene sequences of 355 isolates collected from 1966 to 2019. **(A)** Overview of the spread. **(B)**

Time-calibrated phylogenic tree, the successive periods of dispersal are indicated (tx). **(C-H)** The maximum clade credibility (MCC) tree and the 95% highest posterior density (HPD) regions reflecting the uncertainty related to the phylogeographic inference were mapped. The phylogeographic scenario is displayed with temporal gradients linked to HPD and line colors. Ongoing movements (gradient of colors) are distinguished from completed ones (lines in black). Arrows visualize the direction of the movements. Prints use map data from Mapbox and OpenStreetMap and their data sources. To learn more, visit https://www.mapbox.com/about/maps/ and http://www.openstreetmap.org/copyright.

https://doi.org/10.1371/journal.ppat.1013168.g002

(7) During the last decades of the 20th century, RYMV dispersed eastward and northward from the Kilombero Valley and the Morogoro region, and southward and westward from Lake Victoria. Recently, the virus became widespread in all rice-growing regions with an increase in movements between adjacent localities (Fig 2G and 2H).

Altogether, the phylogeography of RYMV in East Africa reconstructed from the 335 coat protein gene sequences was characterized by these three long-distance dispersal events along a south-north axis from the middle of the 19th century to the second half of the 20th century. In the past decades, a regular short distance dispersal eastward and westward became predominant.

Phylogeography from full-length sequences and ORFs of 50 representative isolates

The spatio-temporal dispersal of RYMV in East Africa was reconstructed from sequences of 50 isolates representative of the genetic and geographic diversity of the virus in East Africa: full-length sequences (4468 nt, but excluding recombinant regions), sequences of ORF2a (including protease and VPg, 459 codons), ORF2b (including polymerase; 540 codons) and ORF4 (capsid protein, 240 codons). The differences in nucleotide diversity and in the selection pressure among the ORFs are given in the S5 Table. No recombination event was detected within the ORF2a, ORF2b and ORF4.

To make up for the weak temporal signal of these small datasets (50 sequences), a temporal prior was applied to the root of each tree by using the TMRCA of the EA335 tree, following a normal distribution with a mean of 170 years and a standard deviation of 25 years. Regardless the differences in sequence length, diversity and selection pressure between the ORFs (S5 Table), the phylogeographic patterns were similar (Fig 3), and consistent with that reconstructed from the 335 sequences of the coat protein gene: (1) an emergence in the Eastern Arc Mountains, (2) a rapid spread to the adjacent large rice growing valley of Kilombero, (3–4) a short range dispersal towards Morogoro and a long range dissemination northward towards Lake Victoria, (5) spread towards the south of Lake Malawi, (6) spread towards northern Ethiopia, (7) generalized dissemination throughout East Africa.

Phylogeography of the three main lineages

The initial diversification of RYMV in East Africa resulted in the emergence of the three lineages S4, S5 and S6 in the mid-19th century within or near the Kilombero valley. The phylogeography of the three lineages is markedly different (Fig 4A and 4B). The S4 lineage had the greatest epidemic success. It was found south of Lake Victoria in the second half of the 19th century (S4lv), and then circulated around Lake Victoria (S4ug) before dispersing northward into Ethiopia (S4et), then southward to Malawi (S4lm) and finally westward to the Republic of Congo, Rwanda and Burundi. The S6 lineage remained confined to the Kilombero valley and in the Morogoro region for several decades. In the last decades, it spread towards eastern Tanzania, south-west Kenya and to the islands of Zanzibar and Pemba. The S7 strain split early from the S6 lineage and was recovered recently in southern Malawi. The S5 lineage was found exclusively in the Kilombero valley and in the Morogoro region. The migration curve (Fig 4C) illustrates the differences in velocity among the lineages, with an early and rapid spread of the S4 lineage, a late but rapid spread of the S6 lineage (except the early spread of the S7 strain), and a lack of dispersal for S5. Interestingly, there was a slowdown in the dispersal rates the past decades within most strains of the S4 and S6 lineages.
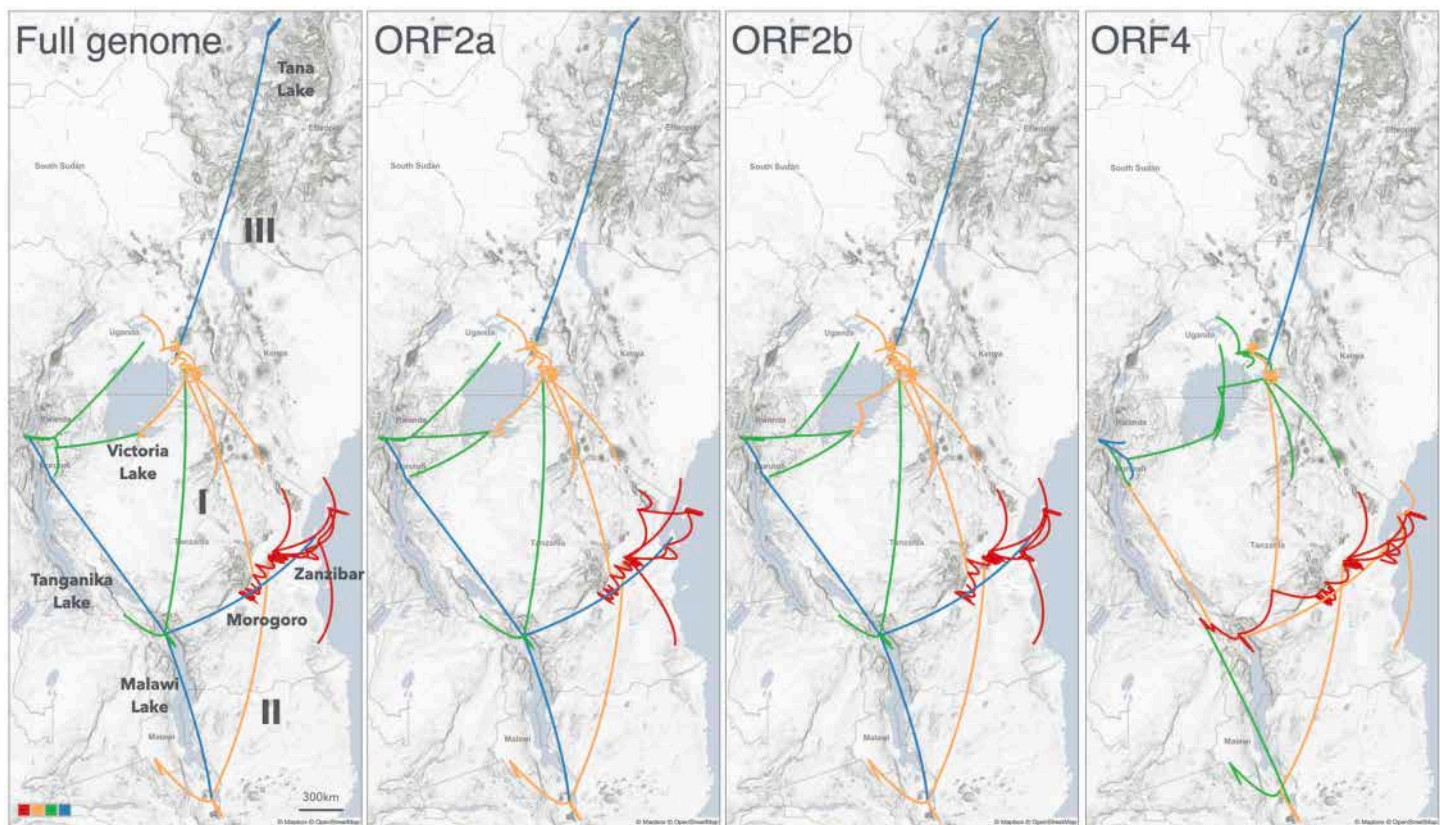
Fig 3. Phylogeographic reconstructions of RYMV in East Africa from the different ORFs. Continuous phylogeographic reconstructions of the spatiotemporal dispersal of RYMV lineages in East Africa from 1850 to 2020 based on the non-recombinant full-length sequences, ORF2a and ORF2b and ORF4 of 50 representative isolates of the genetic and geographic diversity. The recombinant regions were removed from the full-length sequences (see Materials and Methods). The phylogeographic scenario is displayed with temporal gradients linked to line color and curvature. Prints use map data from Mapbox and OpenStreetMap and their data sources. To learn more, visit https://www.mapbox.com/about/maps/ and http://www.openstreetmap.org/copyright.

https://doi.org/10.1371/journal.ppat.1013168.g003

Additional surveys were conducted from 2013 to 2017 at the Kilombero Rice Plantation (KPL) rice intensification station, located within the Kilombero Valley, which covers 5 000 ha of rice. A total of 75 isolates was collected in rice fields under intensifying schemes and in surrounding traditional rice fields. The coat protein gene of these isolates was sequenced. Almost all of the isolates belong to strain S6b already found in this region. Only one isolate of strain S6a, one S5 isolate and four S4lm isolates were collected. The low number of isolates, other than of the S6b strain, shows that the introduction of strains in the Kilomobero rice valley was restricted, even in the present time, reflecting its epidemiological isolation.

East African origin of the introduction of RYMV into West Africa and into Madagascar

The main trends of dispersal of RYMV throughout East Africa, West Africa, Madagascar were captured within a single RRW model. This underlines the flexibility of the RRW models to reconstruct spatio-temporal dispersal when long distance jumps, even transcontinental ones, occurred. The phylogeny of RYMV was reconstructed from the full-length sequences of 101 isolates representative of the genetic and geographic diversity of the virus in Africa: 50 isolates from East Africa, 45 from West Africa and six sequences from Madagascar. A major split in the phylogenetic network separates East Africa and
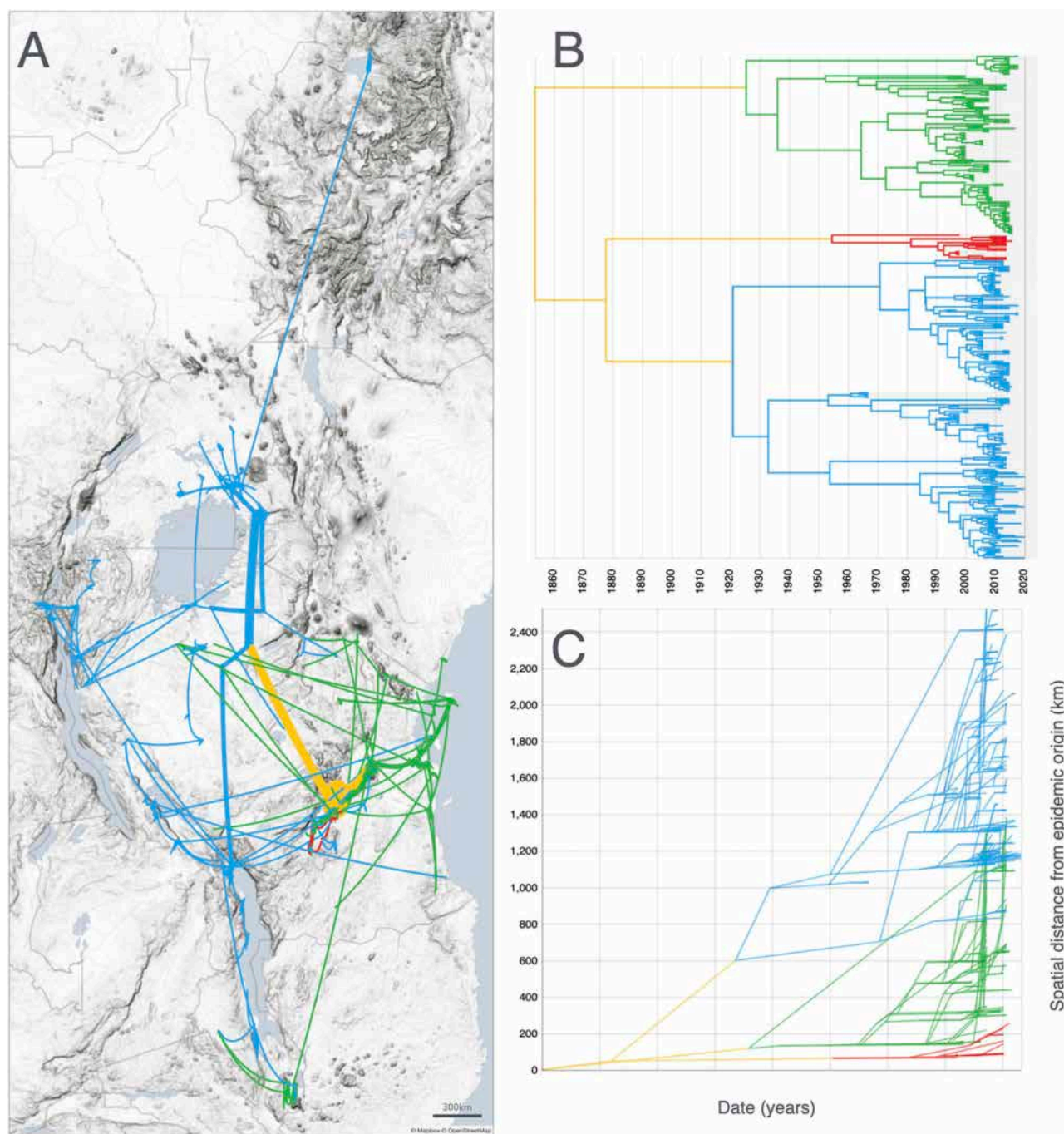
Fig 4. Phylogeography of the three main lineages of RYMV in East Africa. (A) Continuous phylogeographic reconstruction of the spatiotemporal dispersal of the three main lineages of RYMV in East Africa from 1850 to 2020 based on the coat protein gene sequences of 355 isolates collected from 1966 to 2019. The three lineages were distinguished by their colors: S4 (blue), S6 (green), S5 (red). Note that the S5 lineage, confined to the Kilombero valley and adjacent Morogoro region, is nearly masked by the S6 lineage that has a broader distribution. **(B)** Maximum credibility tree with time line reconstructed from the dataset and used to generate the spatio-temporal dispersal (same color code for each lineage). **(C)** Migration curves over time (preorder transversal of the phylogenetic tree) for each parent/child node pair: X-axis = cumulative branch length (as dates in years), Y-axis = cumulative great-circle distances given latitudes and longitudes. Prints use map data from Mapbox and OpenStreetMap and their data sources. To learn more, visit https://www.mapbox.com/about/maps/ and http://www.openstreetmap.org/copyright.

https://doi.org/10.1371/journal.ppat.1013168.g004

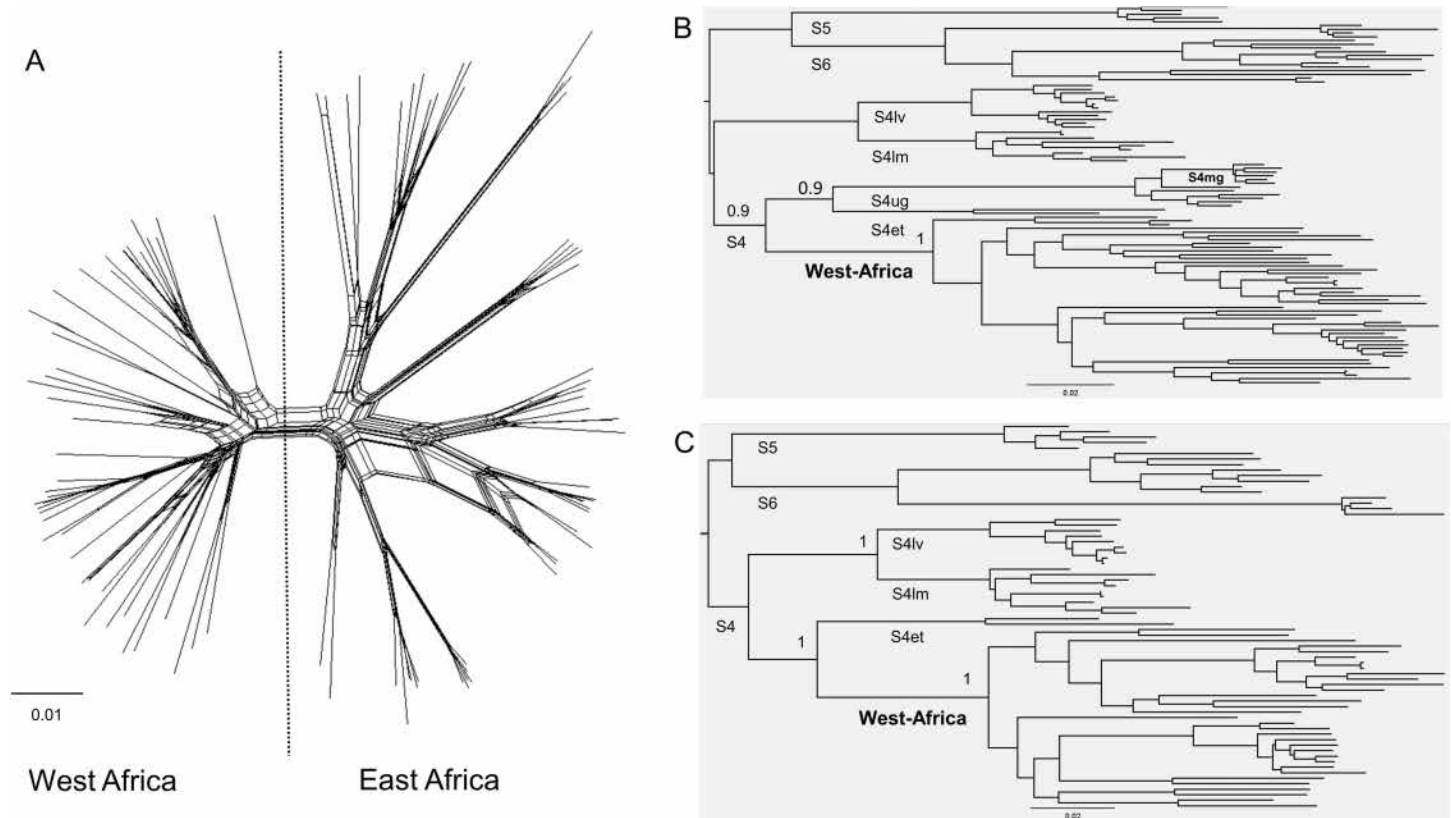**Fig 5. NeighborNet and phylogeny of RYMV in Africa reconstructed from the full genome.** NeighborNet phylogenetic network of the 101 fully sequenced isolates of RYMV from East Africa and West Africa under a HKY85 distance model **(A)**. Maximum likelihood phylogeny tree reconstructed under a HKY85 model and mid-point rooting of the recombination-free full sequences obtained after removing the recombinant regions **(B)**, and after removing the recombinant sequences **(C)**.

https://doi.org/10.1371/journal.ppat.1013168.g005

Madagascar isolates from West Africa ones (Fig 5A). The phylogeny is characterized by the nested position of the West African lineage within the East African lineage comprising the S4ug, S4mg and S4et strains (Fig 5B). The S4mg strain in Madagascar resulted from a recombination between the S4ug strain and an unidentified minor parent [14]. The relationship between the West-African lineage and the S4et strain of this lineage persist after excluding the S4mg and S4ug sequences (Fig 5C). This confirms the earlier assumptions, made on a restricted dataset, that RYMV in West Africa and in Madagascar originated in East Africa [14,15], and brings information on the source lineage of RYMV in West Africa.

We detailed the East African origin of the introduction of RYMV into West Africa and into Madagascar. Convergence of the analyses with the full dataset (AE335, WA335 and Mg94 dataset, i.e., a total of 784 coat protein gene sequences) was difficult to reach. Therefore, 50 isolates representative of the diversity of the virus in West Africa and 15 isolates representative of Madagascar were selected using PDA analyzer [43]. Then we reconstructed the phylogeography of the collection of the coat protein gene of 335 isolates from East Africa, the 50 representative isolates from West Africa and the 15 representative isolates from Madagascar, i.e., a total of 400 isolates. The phylogeographic model was that defined for the analysis of the EA335 sequence dataset (see Materials and Methods). The split between the East and the West African clades occurred in ca. 1874 (95% HDP 1843–1912) at the south of Lake Victoria (Fig 5). The introduction of RYMV in the Niger Inner Delta in West Africa occurred in ca. 1897 (95% HDP 1870–1921). The subsequent pattern of dispersion throughout West Africa and toward Central Africa mirrored the findings from comprehensive surveys conducted in West

Table 2. Dispersal metrics of RYMV in East Africa, West Africa, Madagascar and the three regions combined[1].

Dataset	Weighted diffusion coefficient (km²/year)	IBD signal (rP)[2]	Number of samples	Study area (km²)[3]	Reference
East Africa	1081 [897, 1257]	0.37 [0.30, 0.45]	335	~2 103 000	this study
Madagascar	1246 [622, 1876]	0.35 [0.28, 0.44]	94	~446 000	[14]
West Africa	1603 [1159, 2166]	0.56 [0.45, 0.64]	210	~2 979 000	[74]
West Africa and the Niger Valley	1258 [953, 1585]	0.52 [0.44, 0.62]	261	~3 087 000	[44]
East Africa, West Africa and Madagascar combined	2718 [2259, 3379]	0.39 [0.24, 0.60]	400	~12 522 000	this study

[1]For each dataset and metric, we report both the posterior median estimate and the 95% highest posterior density (HPD) interval.

[2]The isolation-by-distance (IBD) signal has been estimated by the Pearson correlation coefficient (rP) between the patristic and log-transformed great-circle geographic distances computed for each pair of virus samples.

[3]The extent of the study areas were approximated by the inland area of the minimum convex hull polygon surrounding all the sampling.

Africa [44,45]. In Madagascar, RYMV was introduced in ca. 1980 (95% HPD 1975–1986) from the northeast of Lake Victoria. The introduction of RYMV in the northwest of Madagascar followed by the spread of the virus southward of the country is a scenario similar to that obtained from the comprehensive dataset [14].

Comparative dispersal statistics of RYMV

For RYMV, variation in isolation-by-distance (IBD) signals among the three regions (East Africa, West Africa and Madagascar) was apparent (from 0.35 to 0.56) (Table 2). The IBD index was higher in West Africa (0.52 and 0.56 for the two datasets) than in East Africa (0.37) and in Madagascar (0.35) (Table 2). In East Africa, the recurrent long distance human-mediated dispersal of RYMV in a wide range of directions revealed by this study is consistent with a low IBD signal. The higher IBD signal estimated in West Africa may result from the dominant eastward long-distance transmission events establishing more local transmission chains [44,45,46]. The relatively low IBD signal estimated in Madagascar suggested that long-distance human-mediated dispersal, which blurred the relationship between geographical and genetic distances, is also frequent in a country where rice is widely grown. Diffusion coefficients in the three regions ranged from 1100 to 1600 km² per year with overlapping 95% HPD intervals. This indicates a similar dispersal capacity of the virus across the three regions despite differences in IBD signal values, spatial scales, rice history and cultivation, and possibly also in dispersal modes (Fig 6).

The diffusion coefficient of RYMV for East Africa, West Africa and Madagascar combined (covering ca. 12 million km²) was 2700 km²/year, was twice that of each of three regions reflecting the inclusion of the long-distance dispersal from East Africa towards West Africa and towards Madagascar (Table 2).

Discussion

The historical data provided circumstantial evidence that support, explain and complement the interpretation of the phylogeographic reconstructions. The seven periods identified in the phylogeographic reconstructions of RYVM in East Africa are interpreted as followed on the basis of the history of rice since the beginning of the 19th century.

(1) The postulated emergence of RYMV in the Eastern Arc Mountains [47] was confirmed and characterized in more details. The emergence of RYMV dates back to the mid-19th century. It took place near the Great Ruaha Escarpment of the Udzungwa Mountains adjacent to the Kilombero valley and south to Morogoro. Consistently, the insertion-deletion polymorphism originated in isolates of this region. The original host(s) of RYMV were located within the Eastern Afromontane biodiversity hotspot [15,47]. Slash and burn rice cultivation practiced in the
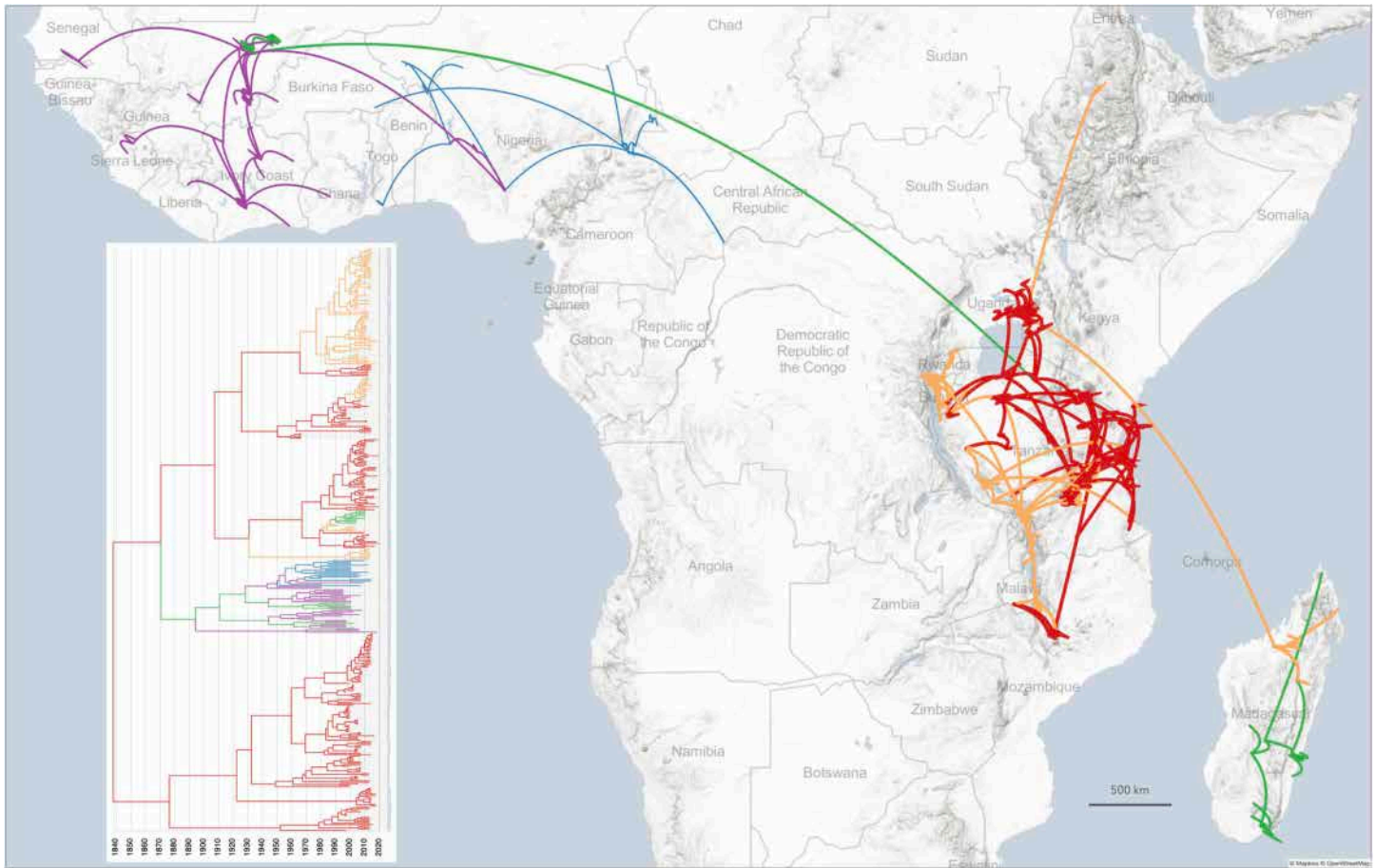
Fig 6. Dispersion of RYMV from East Africa to West Africa and Madagascar. Dispersion of RYMV from East Africa (red and orange) to West Africa (purple and blue) and to Madagascar (pale green) between 1850 and 2020 reconstructed from the ORF4 sequences of 335 isolates from East Africa, 50 isolates from West Africa and 15 isolates from Madagascar. Temporal gradients are linked to line color. Inset: Maximum credibility tree reconstructed from the dataset and used to generate the spatio-temporal dispersal (same color code). Prints use map data from Mapbox and OpenStreetMap and their data sources. To learn more, visit https://www.mapbox.com/about/maps/ and http://www.openstreetmap.org/copyright.

https://doi.org/10.1371/journal.ppat.1013168.g006

mid-1800s near the Great Ruaha Escarpment of the Udzungwa Mountains [29,33,48,49] resulted in spillovers to cultivated rice, presumably – given RYMV host range – from the primary hosts of the *Oryzeae* tribe of the sub-family *Oryzoideae* and of the *Eragrostidae* tribe of the subfamily *Chloridoideae* within the grass family *Poaceae*. Primary hosts may also include the finger millet (*Eleusine corocana*) of the subfamily *Chloridoideae*, a local host of RYMV [5,6], widely cultivated through slash and burn cultivation in these mountains during the 19th century [48]. This led to introductions of RYMV in cultivated rice. The basal split between the S4, S5 and S6 lineages and the insertion-deletion polymorphism suggest (at least) three independent spillovers from the wild host(s) to cultivated rice. Strains specific poaceae species such as *Dactyloctenium aegyptium* have been identified [50]. Specific mutations of the S7 strain occurred at positions of the VPg under positive selection and experimentally associated to host range [51,52]. Further studies are necessary to understand the role of the genetic changes, if any, associated to host jumps and rice spillover.

(2) Within a few decades, RYMV spread to the large adjacent Kilombero Valley where rice had been widely cultivated since the beginning of the 19th century [29]. This is consistent with the movements of rice-growing populations between the Udzungwa Mountains and the Kilombero Valley [29].

(3) Between the second half of the 19th century and the beginning of the 20th century, RYMV was introduced in the Morogoro region and spread to the south of Lake Victoria 400 km northward. Its dispersal followed the introduction of rice into the interior of the continent along the caravan routes linking the coastal strip to Lake Victoria [53, 54, 55]. There are indications that the caravan routes crossed the Morogoro region and the Kilombero valley [33,56]. Interestingly, the migration pathways of turnip mosaic virus (TuMV), a seed transmitted potyvirus, retraced some of the historical trade arteries of the Silk Road [57]. So, both RYMV and TuMV illustrate the link between plant virus dispersal and trade routes.

(4) In the following decades, up to the middle of the 20th century, RYMV dispersed at the east of Lake Victoria where rice had been introduced at the end of the 19th century, along the caravan roads [56] or by boat [58,59]. Interestingly, the first field report of RYMV in East Africa was in 1966 near Kisumu also at the east of Lake Victoria [5].

(5) At the beginning of the 20th century, RYMV was introduced from the Kilombero valley to the south of Lake Malawi. This movement was parallel and concomitant to the movement of military troops at the end of the First World War [60,61]. Feeding the troops was a major issue [62] and rice was the main staple food for soldiers. The Kilombero valley, the Rufiji basin and the Morogoro region were the main rice production areas. Accordingly, RYMV may have been dispersed through the transport of contaminated seeds during the First World War. An alternative hypothesis, linking virus dispersal to the Kilwa-Nyasa south caravan route, from Kiwa 300 km south of Zanzibar to the south of Lake Malawi (formerly named Lake Nyasa), is less likely. First, there is no firm indication that rice had been introduced at the south of Lake Malawi along this caravan route [63,64]. Second, the peak of activity along the Kilwa-Nyasa south caravan route occurred in the second half of the 19th century [63,64] whereas the introduction of RYMV at the south of Lake Malawi occurred at the beginning of the 20th century.

(6) In the second half of the 20th century, RYMV was introduced into northern Ethiopia from the north of Lake Victoria. Large rice intensification schemes were established in the 1960s around Lake Victoria [5] and in the 1980s around Lake Tana in northern Ethiopia [65]. Seed exchanges between these intensification schemes were frequent and may have introduced RYMV in northern Ethiopia. The widespread cultivation of teff (*Eragrostis tef*), a host of RYMV [5], in the Ethiopian highlands may have facilitated the dissemination.

(7) RYMV then spread eastward and northward from the Morogoro region (strain S6a) and from the Kilombero Valley (strain S6b), and southward and westward from the Lake Victoria region (strains S4). RYMV finally became widespread throughout East Africa. This spread accompanied the expansion of rice cultivation of the last decades of the 20th century and of the beginning of the 21st century. The slowdown in the dispersal rates of several strains of the S4 and S6 lineages over the past decades reflected a shift from an expansion phase to an endemic phase [34,35] in RYMV epidemiology. This may be due to changes in the trade and market patterns [66] with a gradual change from long distance to local seed transport as rice became widely cultivated.

The phylogeography of RYMV throughout East Africa unveiled the role of rice transport in the long-distance movement of the virus. Similarly, the introduction of RYMV in West Africa may result from seed movements. Seed exchanges in Africa between agricultural stations, experimental farms, botanical and missionary gardens, even far apart, were frequent at the end of the 19th century [67,68]. The Niger Inner Delta was a major destination for seeds from a wide range of origins, including Madagascar, because of rice intensification in this region since the end of the 19th century [69,70]. The origin of RYMV introduced in West Africa was located at the south of Victoria Lake at the end of the 19th century, in Tanzania, which was under German occupation at that time. Early contacts between German scientists and agricultural stations

of the Niger Inner Delta are documented [68]. RYMV may have been introduced directly from this region, or indirectly through the Amani research station which was the main agricultural station in East Africa [71]. Our results also confirm the Ugandan origin of the Madagascar strain [14]. Increased seed exchanges between Madagascar and East African countries have occurred during the recent decades [73]. Transport of contaminated rice seeds may also account for this introduction.

Given a host range that is restricted to rice and closely related poaceae species, a direct connection between RYMV epidemiology and rice ecology was anticipated [10]. However, elucidating the precise nature of this relationship has proven challenging. Linear relationships between the intensity of rice cultivation and the transition frequency of RYMV lineages in West-Africa have been proposed [15], as well as between rice cultivation intensity and virus population size in the irrigated valley of Niger [44]. Nevertheless, further statistical analyses failed to provide supports for such associations between the dispersal dynamic of viral lineages or virus population size and rice-cultivated areas [44,74]. Factors that make the relationship between RYMV epidemiology and its hosts more complex than initially proposed in early modelling attempts have been explored [14,44]. Interestingly, in East Africa, the transfer of rice seeds, rather than rice cultivation was critical in RYMV dispersal. Virus spread in East Africa was more closely associated with human-mediated long-distance transport of rice seeds than with the short-range vector-borne movements typically linked to rice cultivation.

RYMV dispersal through the transport of contaminated seeds shows the importance of this neglected virus transmission pathway in plant virus epidemiology [1]. This study also unveils the wide range of human activities, some of them unsuspected, involved in man-mediated movements of a plant virus. The diffusion coefficient of RYMV was higher than that of several animal viruses including rabies in various hosts, even when occasional rapid long-distance dispersal events driven by human-mediated movements occurred [35]. Yet, the diffusion coefficient of RYMV was lower by an order of magnitude than that of animal viruses with long distance dispersal through animal trade or through migratory birds [35].

Maize streak virus A (MSV-A) causes another major cereal disease in Africa [75]. It is transmitted by leafhoppers and not by seeds. Interestingly, introductions of MSV-A from Uganda to Madagascar [73] and from Kenya to Ethiopia [76] have been reported. These movements, parallels to those of RYMV, were also attributed to human-mediated introductions [73,76]. It shows that the spread of RYMV and MSV-A, two major vector borne viruses, does not necessarily, or even primarily, depend on vector transmission.

An important pending question is why RYMV, given its wide dispersal capacities, is still restricted to Africa and does not have a worldwide distribution, unlike several sobemoviruses with similar biological properties. Long-distance movement of RYMV across the sea is possible but is quite a rare event as indicated by the unique successful introduction of the virus in West Africa. Rice was imported into America from West Africa in the 17th century and early 18th centuries, at least a century before the introduction of RYMV in West Africa. In the 19th century, rice exports from East Africa were limited and directed towards regions where rice was not cultivated [31]. In the 20th century, rice was imported into Africa rather than exported from this continent. Altogether, this could explain why RYMV remained restricted to Africa. However, the risk of accidental introduction of RYMV into other continents subsists, in particular through the transfer of seeds for agronomical breeding which we have found to be a decisive means of long distance dissemination of the virus in Africa. Incidentally, MSV confined so far to Africa has recently been reported in the Philippines [77], a possible illustration of such human-mediated trans-continental introductions of vector borne plant viruses.

Materials and methods

Datasets

A collection of 335 capsid protein gene sequences of East African isolates was built. The dataset consists of 240 sequences retrieved from NCBI (2024-06-01) supplemented by 95 sequences obtained in this study (S1 Table). The isolates originated from eight countries (Burundi, Republic of Congo, Ethiopia, Kenya, Malawi, Rwanda, Tanzania, Uganda). The sampling locations were distributed between latitudes 15.6° South and 11.9° North, and from longitudes 28.8° West to 39.8° East, ca.

3 000 kilometers in latitude x 1 200 km in longitude, i.e., a study area of ca. two million square kilometers. The isolates were sampled from 1966 to 2019 (54 years). A minor variant named S4ke recently identified in Kenya with a recombinant event in the capsid protein [78] was considered in this study (see below) but not included in this corpus.

The genome of RYMV is organized into five overlapping open reading frames. ORF1, located at the 5' end, encodes a protein involved in virus movement and gene silencing suppression. ORFx overlaps with the 5' end of ORF2a and is translated through a leaky scanning mechanism. ORF2a and ORF2b are overlapping ORFs that encode the central polyprotein. ORF2a encodes a serine protease and the VPg. ORF2b translated through a -1 ribosomal frameshift encodes the RNA-dependent RNA polymerase. ORF4 located at the 3' end encodes the coat protein. This ORF is expressed via a subgenomic RNA [3].

A collection of full-length sequences of 101 isolates from Africa was set. This dataset consisted of sequences of 50 isolates of East Africa and of the East of Central Africa (referred to as the East African dataset), 45 isolates of West Africa and of the West of Central Africa (referred to as the West African dataset) and six sequences of Madagascar. Among the 50 full-length sequences of the East African dataset, 23 were sequenced for this study and 27 were retrieved from NCBI (S2 Table). This sample is representative of the genetic and geographic diversity of RYMV in East Africa, including the S4ke variant from western Uganda. No recombination signal was detected within the ORF2a and ORF2b genes.

In complement, a collection of 335 capsid protein gene sequences from West African isolates, was considered to estimate the substitution rate of RYMV (S3 Table). It consists of sequences retrieved from NCBI (2024-06-01), 261 sequences used in a study of the phylogeography in West Africa [44] complemented by 46 sequences of Ghana [45] and 28 sequences of Burkina Faso [46]. The 28 sequences from Burkina Faso were chosen from the 144 available on NCBI using the PDA software [43] in order to maximize the genetic diversity of the sample. In total, the samples were collected in 15 countries (Benin, Burkina Faso, Cameroon, Central African Republic, Chad, Gambia, Ghana, Guinea, Ivory Coast, Mali, Niger, Nigeria, Senegal, Sierra Leone, Togo). They were sampled between 1975 and 2021 (47 years).

Phylogenetic reconstructions

Sequences were aligned using CLUSTAL X with default parameters [79]. Maximum-likelihood (ML) phylogenetic trees were inferred using the PHYML-3.1 algorithm implemented in the SEAVIEW version 4.7 software [80] under a HKY85 substitution model. The ML trees were rooted at the point in the tree that minimizes the residual mean square of the root-to-tip distances. The neighborNet phylogenetic network of the 101 fully sequenced isolates was inferred [81] under an HKY85 distance model as implemented in SplitsTree version 4. The neighborNet phylogenetic network illustrates the genetic relationships between sequences taking into account the conflicting phylogenetic signals that are possibly due to recombination events. The full-length sequence alignment was screened for recombination signals using RDP5 version [82]. The default settings were used for each of the seven recombination detection algorithms that RDP incorporates, as was a Bonferroni corrected P-value cut-off of 0.05. Only recombination events detected by more than four of the seven methods implemented in RDP were considered. The phylogeny was reconstructed from recombinant-free alignments. This was obtained by removing either the recombinant sequences or the recombinant regions. In the latter case, it was checked that further recombination was excluded through the pairwise homoplasy (PHI) test [83] implemented in SplitsTree version 4 [81]. Total, synonymous, and non-synonymous nucleotide diversity of ORF2a, ORF2b and ORF4 were assessed by DnaSP version 6 [84]. The ORF1, too short, too variable and with a low phylogenetic signal [85], was not considered.

Bayesian evolutionary inferences

RYMV is a measurably evolving population, yet its overall temporal signal, depending on the dataset, is weak to moderate [14,15]. The temporal signal of each data set was assessed through tip cluster-randomization tests in BEAST [86] and by BETS [37] under a Bayesian statistical framework, and by TreeTime under a likelihood framework [87]. We subsequently reconstructed time-calibrated phylogenies using a Bayesian statistical framework implemented in BEAST version 1.10.4

software [88]. BEAST uses Markov Chain Monte Carlo integration to average over all plausible evolutionary histories for the data, as reflected by the posterior probability. All analyses were performed using the BEAGLE library to enhance computation speed [89,90]. We specified an HKY85 substitution model with a discretized gamma distribution (four categories) to model rate heterogeneity across sites. To accommodate variation of substitution rates across lineages, an uncorrelated relaxed molecular clock that models the branch rate variation according to a lognormal distribution was specified [91]. The flexible nonparametric demographic skygrid prior was selected to accommodate for variation in the rate of coalescence through time [92]. Stationarity and mixing (e.g., based on effective sample sizes > 200 for the continuous parameters) were assessed using Tracer version 1.7 [93]. Consensus trees deriving from the phylogenies sampled during the MCMC analysis were obtained using TreeAnnotator version 1.10.4 [88].

To study the geographical spread of RYMV in continuous space in East Africa and to quantify its tempo and dispersal, we fitted a continuous phylogenetic diffusion model to the East African dataset, modelling the changes in coordinates (latitude and longitude) along each branch in the evolutionary history as a bivariate normal random deviate [22]. As a more realistic alternative to homogeneous Brownian motion, we adopted a "Cauchy" RRW extension. It models variation in dispersal rates across branches by independently drawing branch-specific rate scalers of the RRW precision matrix from a gamma distribution with shape and scale equal to 1/2 to relax the assumption of a constant spatial diffusion rate throughout the whole tree [22]. The RRW that was used in this study is equivalent to a Cauchy process on a tree with rescaled branch lengths, which is a pure jump process allowing for long-distance dispersion events [72]. Bayesian inference using continuous diffusion models yields a posterior distribution of the phylogenetic trees, each having ancestral nodes annotated with location estimates. The EA335 dataset was used to reconstruct the dispersal of RYMV throughout East Africa. A spatial jitter of 0.1° in latitude and in longitude (i.e., ca 10 km) was applied to the locations of the isolates. This degree of noise for identical coordinates was needed to avoid improper posteriors under the RRW model and associated inference problems [22,94]. Evolaps version 3 was used to visualize the continuous phylogeographic reconstruction of the dispersal history of RYMV lineages [95,96]. The phylogeography of RYMV in East Africa was reconstructed from the coat protein gene sequences of the full dataset (335 isolates), from the full genome and different ORFs of a subsample of the 50 isolates representative of the genetic and geographic diversity of the virus.

Comparative dispersal statistics of RYMV

The phylogeography of RYMV has been reconstructed in East Africa (this study), in West Africa [44] and in Madagascar [14]. The dispersal capacity and dynamics in the three regions was compared through the weighted diffusion coefficient (km^2/year) [15] and the isolation-by-distance (IBD) signal metric, two metrics identified as robust (to sampling intensity in particular) and measuring complementary aspects of the overall dispersal pattern [35]. The diffusion coefficient is an estimate of the dispersal capacity of the lineages. The IBD signal metric quantifies how the dispersal is spatially structured, i.e., the tendency of phylogenetically closely related tip nodes to be sampled from geographically close locations. An IBD signal was apparent through the relationships between geographic and genetic pairwise distances in a restricted dataset of sequences from East and West Africa [97]. The diffusion coefficient of RYMV virus over the three regions combined was also estimated. Then, the dispersal capacity of this plant virus was compared to that of a range of animal viruses in various host species [35].

Supporting information

S1 Table. Name and accession number of the 335 isolates from East Africa.
(XLSX)

S2 Table. Name and accession number of the 101 fully sequenced isolates from West Africa, East Africa and Madagascar.
(XLSX)

S3 Table. Name and accession number of the 335 isolates from West Africa.
(XLSX)

S4 Table. Name and accession number of the 94 isolates of Madagascar.
(XLSX)

S5 Table. Length, diversity and selection pressure of the ORF of Rice yellow mottle virus.
(XLSX)

Acknowledgments

Author contributions

Conceptualization: Pauline Rocu, Simon Dellicour, Philippe Lemey, Erik Gilbert, Marie-José Dugué, François Chevenet, Paul Bastide, Stéphane Guindon, Denis Fargette, Eugénie Hébrard.

Data curation: Agnès Pinel-Galzi.

Investigation: Innocent Ndikumana, Geoffrey Onaga, Agnès Pinel-Galzi, Pauline Rocu, Judith Hubert, Hassan Karakacha Wéré, Antony Adego, Mariam Nyongesa Wéré, Simon Dellicour, François Chevenet, Paul Bastide, Denis Fargette.

Project administration: Denis Fargette, Eugénie Hébrard.

Resources: Innocent Ndikumana, Geoffrey Onaga, Judith Hubert, Hassan Karakacha Wéré, Antony Adego, Mariam Nyongesa Wéré.

Software: Maxime Hebrard, François Chevenet.

Writing – original draft: Denis Fargette.

Writing – review & editing: Innocent Ndikumana, Geoffrey Onaga, Agnès Pinel-Galzi, Pauline Rocu, Judith Hubert, Hassan Karakacha Wéré, Antony Adego, Mariam Nyongesa Wéré, Nils Poulicard, Maxime Hebrard, Simon Dellicour, Philippe Lemey, Erik Gilbert, Marie-José Dugué, François Chevenet, Paul Bastide, Stéphane Guindon, Denis Fargette, Eugénie Hébrard.

References

1. Jones RAC. Plant and insect viruses in managed and natural environments: novel and neglected transmission pathways. Adv Virus Res. 2018;101:149–87. https://doi.org/10.1016/bs.aivir.2018.02.006 PMID: 29908589

2. Peters D, Matsumura EE, van Vredendaal P, van der Vlugt RAA. The plant virus transmissions database. J Gen Virol. 2024;105(3):001957. https://doi.org/10.1099/jgv.0.001957 PMID: 38441560

3. Sõmera M, Fargette D, Hébrard E, Sarmiento C, Ictv Report Consortium. ICTV virus taxonomy profile: solemoviridae 2021. J Gen Virol. 2021;102(12):001707. https://doi.org/10.1099/jgv.0.001707 PMID: 34951396

4. Abo M, Sy A, Alegbejo M. Rice yellow mottle virus in Africa: evolution, distribution, economic significance and sustainable rice production and management strategies. J Sustain Agric. 1997;11:85–111.

5. Bakker W. Characterization and ecological aspects of rice yellow mottle virus in Kenya. Agri Res Rep. 1974;829:1–152.

6. Allarangaye MD, Traoré O, Traoré EVS, Millogo RJ, Konaté G. Evidence of non-transmission of rice yellow mottle virus through seeds of wild host species. J Plant Pathol. 2006;88:309–15.

7. Fauquet C, Thouvenel J. Isolation of the rice yellow mottle virus in Ivory Coast. Plant Dis Rep. 1977;61:443–6.

8. Konaté G, Sarra S, Traoré O. Rice yellow mottle virus is seed-borne but not seed transmitted in rice seeds. Eur J Plant Pathol. 2001;107:361–4.

9. Sarra S, Peters D. Rice yellow mottle virus is transmitted by cows, donkeys, and grass rats in irrigated rice crops. Plant Dis. 2003;87(7):804–8. https://doi.org/10.1094/PDIS.2003.87.7.804 PMID: 30812890

10. Traoré O, Pinel-Galzi A, Sorho F, Sarra S, Rakotomalala M, Sangu E, et al. A reassessment of the epidemiology of Rice yellow mottle virus following recent advances in field and molecular studies. Virus Res. 2009;141(2):258–67. https://doi.org/10.1016/j.virusres.2009.01.011 PMID: 19195488

11. Cooper JI, MacCallum F. Viruses and the Environment. 1st eds. Chapman and Hall, Springer Netherlands; 1984.

12. Teakle D. Abiotic transmission of southern bean-mosaic virus in soil. Aust J Biol Sci. 1986;39:353–60.

13. Traoré O, Traoré MD, Fargette D, Konaté G. Rice seedbeds as a source of primary infection by Rice yellow mottle virus. Eur J Plant Pathol. 2006;115(2):181–6. https://doi.org/10.1007/s10658-006-9004-9

14. Rakotomalala M, Vrancken B, Pinel-Galzi A, Ramavovololona P, Hébrard E, Randrianangaly JS, et al. Comparing patterns and scales of plant virus phylogeography: Rice yellow mottle virus in Madagascar and in continental Africa. Virus Evol. 2019;5(2):vez023. https://doi.org/10.1093/ve/vez023 PMID: 31384483

15. Trovão NS, Baele G, Vrancken B, Bielejec F, Suchard MA, Fargette D, et al. Host ecology determines the dispersal patterns of a plant virus. Virus Evol. 2015;1(1):vev016. https://doi.org/10.1093/ve/vev016 PMID: 27774287

16. Abé Y. Le "décorticage" du riz. 1st ed. ed. Maison des Sciences de l'Homme; 2007.

17. Banwo OO, Alegbejo MD, Abo ME. Rice yellow mottle virus genus Sobemovirus: a continental problem in Africa. Plant Prot Sci. 2004;40(1):26–36. https://doi.org/10.17221/1350-pps

18. Richards P. Coping with hunger: hazard and experiment in an African rice-farming system. 1st ed. Allen and Unwin, Routledge. 1986.

19. Pedgley D. Windborne pests and diseases. 1st ed. ed. Horwood E, UK: John Wiley and Sons. 1982.

20. Thresh J. The long-range dispersal of plant viruses by arthropod vectors. Philos Trans R Soc Lond B Biol Sci. 1983;302:497–528.

21. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009;5(9):e1000520. https://doi.org/10.1371/journal.pcbi.1000520 PMID: 19779555

22. Lemey P, Rambaut A, Welch JJ, Suchard MA. Phylogeography takes a relaxed random walk in continuous space and time. Mol Biol Evol. 2010;27(8):1877–85. https://doi.org/10.1093/molbev/msq067 PMID: 20203288

23. Fargette D, Pinel A, Rakotomalala M, Sangu E, Traoré O, Sérémé D, et al. Rice yellow mottle virus, an RNA plant virus, evolves as rapidly as most RNA animal viruses. J Virol. 2008;82(7):3584–9. https://doi.org/10.1128/JVI.02506-07 PMID: 18199644

24. Ghafari M, Sõmera M, Sarmiento C, Niehl A, Hébrard E, Tsoleridis T, et al. Revisiting the origins of the Sobemovirus genus: a case for ancient origins of plant viruses. PLoS Pathog. 2024;20(1):e1011911. https://doi.org/10.1371/journal.ppat.1011911 PMID: 38206964

25. Carpenter A. The history of rice in Africa. In: Buddenhagen IW, Persley GJ, editors. Rice in Africa. Academic Press; 1978. p. 3–10.

26. Rice Gilbert E.civilization and the swahili towns: anti-commodity and anti-state?. In: Hazareesingh S, Maat H, editors. Local subversions of colonial cultures. London: Palgrave Macmillan; 2016. p. 170–86.

27. Portères R. Riz. In: Bertin J, Hémardinquer J, Keul M, Randles W, editors. Atlas des cultures vivrières. Paris: Ecole Pratique des Hautes Etudes. 1971. p. 4.

28. Baur R, Le Roy A. A travers le Zanguebar. 1st ed. Alfred Mame et Fils. 1884.

29. Jatzold R, Baum E. The Kilombero valley. Characteristic features of the economic geography of a semihumid East African flood plain and its margins. Weltforum Verlag. 1968.

30. Alpers EA. The Western Indian Ocean as a regional food network in the nineteenth century. International Congress on Indian Ocean studies, Perth, Western Australia, 5-12 December 1984. Proceedings Section A: Resources, Environment and Economic Development, 20 pp.

31. Alpers EA. East Africa and the Indian Ocean. Princeton: Markus Wiener Publishers; 2009.

32. Beardall W. Exploration of the Rufiji river under the orders of the Sultan of Zanzibar. Proc R Geogr Soc. 1881;II(11):641–56.

33. Iliffe J. A modern history of Tanganyika. Cambridge University Press. 1979.

34. Bastide P, Rocu P, Wirtz J, Hassler GW, Chevenet F, Fargette D, et al. Modeling the velocity of evolving lineages and predicting dispersal patterns. Proc Natl Acad Sci U S A. 2024;121(47):e2411582121. https://doi.org/10.1073/pnas.2411582121 PMID: 39546571

35. Dellicour S, Bastide P, Rocu P, Fargette D, Hardy OJ, Suchard MA, et al. How fast are viruses spreading in the wild?. PLoS Biol. 2024;22(12):e3002914. https://doi.org/10.1371/journal.pbio.3002914 PMID: 39625970

36. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, et al. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. Proc Natl Acad Sci U S A. 2012;109(37):15066–71. https://doi.org/10.1073/pnas.1206598109 PMID: 22927414

37. Duchene S, Lemey P, Stadler T, Ho SYW, Duchene DA, Dhanasekaran V, et al. Bayesian evaluation of temporal signal in measurably evolving populations. Mol Biol Evol. 2020;37(11):3363–79. https://doi.org/10.1093/molbev/msaa163 PMID: 32895707

38. Pybus OG, Rambaut A, Belshaw R, Freckleton RP, Drummond AJ, Holmes EC. Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. Mol Biol Evol. 2007;24(3):845–52. https://doi.org/10.1093/molbev/msm001 PMID: 17218639

39. Mbewe W, Mukasa S, Ochwo-Ssemakula M, Sseruwagi P, Tairo F, Ndunguru J, et al. Cassava brown streak virus evolves with a nucleotide-substitution rate that is typical for the family Potyviridae. Virus Res. 2024;346:199397. https://doi.org/10.1016/j.virusres.2024.199397 PMID: 38750679

40. Fargette D, Pinel A, Abubakar Z, Traoré O, Brugidou C, Fatogoma S, et al. Inferring the evolutionary history of rice yellow mottle virus from genomic, phylogenetic, and phylogeographic studies. J Virol. 2004;78(7):3252–61. https://doi.org/10.1128/jvi.78.7.3252-3261.2004 PMID: 15016846

41. Guindon S, De Maio N. Accounting for spatial sampling patterns in Bayesian phylogeography. Proc Natl Acad Sci U S A. 2021;118(52):e2105273118. https://doi.org/10.1073/pnas.2105273118 PMID: 34930835

42. Kalkauskas A, Perron U, Sun Y, Goldman N, Baele G, Guindon S, et al. Sampling bias and model choice in continuous phylogeography: Getting lost on a random walk. PLoS Comput Biol. 2021;17(1):e1008561. https://doi.org/10.1371/journal.pcbi.1008561 PMID: 33406072

43. Chernomor O, Minh BQ, Forest F, Klaere S, Ingram T, Henzinger M, et al. Split diversity in constrained conservation prioritization using integer linear programming. Methods Ecol Evol. 2015;6(1):83–91. https://doi.org/10.1111/2041-210X.12299 PMID: 25893087

44. Issaka S, Traoré O, Longué RDS, Pinel-Galzi A, Gill MS, Dellicour S, et al. Rivers and landscape ecology of a plant virus, Rice yellow mottle virus along the Niger Valley. Virus Evol. 2021;7(2):veab072. https://doi.org/10.1093/ve/veab072 PMID: 36819970

45. Omiat E, Asante M, Traoré V, Oppong A, Ifie B, Ofosu K, et al. Genetic diversity and epidemic histories of rice yellow mottle virus in Ghana. Viruses. 2023;329:199106.

46. Billard E, Barro M, Sérémé D, Bangratz M, Wonni I, Koala M, et al. Dynamics of the rice yellow mottle disease in western Burkina Faso: Epidemic monitoring, spatio-temporal variation of viral diversity and pathogenicity in a disease hotspot. Virus Evol. 2023;9:2, https://doi.org/10.1093/ve/vead049

47. Pinel-Galzi A, Traoré O, Séré Y, Hébrard E, Fargette D. The biogeography of viral emergence: rice yellow mottle virus as a case study. Curr Opin Virol. 2015;10:7–13. https://doi.org/10.1016/j.coviro.2014.12.002 PMID: 25544357

48. Paul JL. Anthropologie historique des hautes terres de Tanzanie orientale. Karthala. 2003.

49. Raison J. Tanzanie: l'Ujamaa et ses lendemains. Les Afriques au sud du Sahara. Paris: Belin-Reclus. 1994. p. 372–88.

50. Allarangaye M, Traoré O, Traoré E, Millogo R, Guinko S, Konaté G. Host range of rice yellow mottle virus in sudano-sahelian savannahs. Pak J Biol Sci. 2007;9:1414–21.

51. Ndikumana I, Pinel-Galzi A, Fargette D, Hébrard E. Complete genome sequence of a new strain of Rice yellow mottle virus from Malawi, characterized by a recombinant VPg protein. Genome Announc. 2017;5(44):e01198-17. https://doi.org/10.1128/genomeA.01198-17 PMID: 29097464

52. Poulicard N, Pinel-Galzi A, Traoré O, Vignols F, Ghesquière A, Konaté G, et al. Historical contingencies modulate the adaptability of Rice yellow mottle virus. PLoS Pathog. 2012;8(1):e1002482. https://doi.org/10.1371/journal.ppat.1002482 PMID: 22291591

53. Finch J, Marchant R, Mustaphi C. Ecosystem change in the South Pare Mountain bloc, Eastern Arc Mountains of Tanzania. Holocene. 2017;27(6). https://doi.org/10.1177/0959683616675

54. Roberts A. Nyamwezi trade. In: Gry R, Birmingham D, editors. Pre-colonial African trade. Essays on trade in central and eastern Africa before 1900. Oxford University Press; 1970.

55. Wynne-Jones S, Croucher S. The central caravan route of Tanzania: a preliminary archaeological reconnaissance. Nyame Akuma. 2007;67:91–5.

56. Koponen J. People and Production in Late Precolonial Tanzania. Uppsala: Scandinavian Institute of African Studies; 1988

57. Kawakubo S, Gao F, Li S, Tan Z, Huang Y-K, Adkar-Purushothama CR, et al. Genomic analysis of the brassica pathogen turnip mosaic potyvirus reveals its spread along the former trade routes of the Silk Road. Proc Natl Acad Sci U S A. 2021;118(12):e2021221118. https://doi.org/10.1073/pnas.2021221118 PMID: 33741737

58. Leclercq J. Aux sources du Nil par le chemin de fer de l'Ouganda. 1913.

59. Gry R, Birmingham D. The northern interlacustrine region. Pre-colonial African trade. Essays on trade in central and eastern africa before 1900. Oxford University Press; 1970.

60. Bovill EW. Notes from East Africa: The Uluguru Mountains and the Rufiji Plain. The Geographical Journal. 1917;50(4):277. https://doi.org/10.2307/1780332

61. Von Lettow-Vorbec P. My reminiscences of East Africa. Pantianos Classics. 1920.

62. Anderson R. Norforce: Major general Edward Northey and the Nyasaland and northeastern Rhodesia frontier force, January 1916 to June 1918. Scientia Militaria. 2016;44(1):47–80.

63. Alpers EA. Ivory and slaves in east central Africa. Changing patterns of international trade to the later nineteenth century. Heinemann. 1975.

64. Biginagwa T, Mapunda B. The Kilwa–Nyasa caravan route the long-neglected trading corridor in southern Tanzania. In: Wynnes-Jones S, LaViolette A, editors. The swahili world. 2017. p. 541–54.

65. Alemu D, Tesfaye A, Assaye A, Addis D, Tadesse T, Thompson J. A historical analysis of rice commercialisation in Ethiopia: the case of the Fogera plain. APRA Working Paper 18. 2018. Future Agricultures Consortium, Brighton. https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/14283

66. Meillassoux C. The development of indigenous trade and markets in West Africa. Routledge; 1971.

67. Bonneuil C. Mettre en ordre et discipliner les tropiques: les sciences du végétal dans l'Empire français, 1870-1940. Université de Paris VII. 1997.

68. Tourte R. La période coloniale et les grands moments des jardins d'essais. 2019.

69. Viguier P. La Riziculture Indigène au Soudan Français. Paris: Larose; 1939.

70. Portères R. Vieilles agricultures de l'Afrique intertropicale. L'Agronomie Tropicale. 1950;5(9–10):489–507.

71. Blais H. L'empire de la nature: une histoire des jardins botaniques coloniaux (fin XVIIIe siècle–années 1930). Champ Vallon Editions; 2023.

72. Bastide P, Didier G. The Cauchy process on phylogenies: a tractable model for pulsed evolution. Syst Biol. 2023;72(6):1296–315. https://doi.org/10.1093/sysbio/syad053 PMID: 37603537

73. Oyeniran KA, Martin DP, Lett J-M, Rakotomalala MR, Azali HA, Varsani A. Movement of the A-strain maize streak virus in and out of Madagascar. Virology. 2024;600:110222. https://doi.org/10.1016/j.virol.2024.110222 PMID: 39265447

74. Dellicour S, Vrancken B, Trovão NS, Fargette D, Lemey P. On the importance of negative controls in viral landscape phylogeography. Virus Evol. 2018;4(2):vey023. https://doi.org/10.1093/ve/vey023 PMID: 30151241

75. Shepherd DN, Martin DP, Van Der Walt E, Dent K, Varsani A, Rybicki EP. Maize streak virus: an old and complex "emerging" pathogen. Mol Plant Pathol. 2010;11(1):1–12. https://doi.org/10.1111/j.1364-3703.2009.00568.x PMID: 20078771

76. Ketsela D, Oyeniran KA, Feyissa B, Fontenele RS, Kraberger S, Varsani A. Molecular identification and phylogenetic characterization of A-strain isolates of maize streak virus from western Ethiopia. Arch Virol. 2022;167(12):2753–9. https://doi.org/10.1007/s00705-022-05614-4 PMID: 36169719

77. Santa Cruz FC, Domalanta JC, Minguez LT, Nieto FGM, Alviar KB. First report of maize streak virus affecting maize in the Philippines. New Disease Reports. 2024;50(1). https://doi.org/10.1002/ndr2.12302

78. Adego AK, Poulicard N, Pinel-Galzi A, Mukoye B, Fargette D, Wéré HK, et al. Full-length genome sequences of recombinant and nonrecombinant sympatric strains of rice yellow mottle virus from western Kenya. Genome Announc. 2018;6(8):e01508-17. https://doi.org/10.1128/genomeA.01508-17 PMID: 29472342

79. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22(22):4673–80. https://doi.org/10.1093/nar/22.22.4673 PMID: 7984417

80. Gouy M, Guindon S, Gascuel O. Seaview version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010;27(2):221–4. https://doi.org/10.1093/molbev/msp259 PMID: 19854763

81. Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol. 2004;21(2):255–65. https://doi.org/10.1093/molbev/msh018 PMID: 14660700

82. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, et al. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. Virus Evol. 2020;7(1):veaa087. https://doi.org/10.1093/ve/veaa087 PMID: 33936774

83. Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. Genetics. 2006;172(4):2665–81. https://doi.org/10.1534/genetics.105.048975 PMID: 16489234

84. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34(12):3299–302. https://doi.org/10.1093/molbev/msx248 PMID: 29029172

85. Pinel-Galzi A, Mpunami A, Sangu E, Rakotomalala M, Traoré O, Sérémé D, et al. Recombination, selection and clock-like evolution of Rice yellow mottle virus. Virology. 2009;394(1):164–72. https://doi.org/10.1016/j.virol.2009.08.008 PMID: 19740507

86. Murray GGR, Wang F, Harrison EM, Paterson GK, Mather AE, Harris SR, et al. The effect of genetic structure on molecular dating and tests for temporal signal. Methods Ecol Evol. 2016;7(1):80–9. https://doi.org/10.1111/2041-210X.12466 PMID: 27110344

87. Sagulenko P, Puller V, Neher RA. TreeTime: maximum-likelihood phylodynamic analysis. Virus Evol. 2018;4(1):vex042. https://doi.org/10.1093/ve/vex042 PMID: 29340210

88. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018;4(1):vey016. https://doi.org/10.1093/ve/vey016 PMID: 29942656

89. Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, et al. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Syst Biol. 2012;61(1):170–3. https://doi.org/10.1093/sysbio/syr100 PMID: 21963610

90. Suchard MA, Rambaut A. Many-core algorithms for statistical phylogenetics. Bioinformatics. 2009;25(11):1370–6. https://doi.org/10.1093/bioinformatics/btp244 PMID: 19369496

91. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006;4(5):e88. https://doi.org/10.1371/journal.pbio.0040088 PMID: 16683862

92. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol Biol Evol. 2013;30(3):713–24. https://doi.org/10.1093/molbev/mss265 PMID: 23180580

93. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using tracer 1.7. Syst Biol. 2018;67(5):901–4. https://doi.org/10.1093/sysbio/syy032 PMID: 29718447

94. Fernández C, Steel MFJ. Bayesian regression analysis with scale mixtures of normals. Econom Theory. 2000;16(1):80–101. https://doi.org/10.1017/s0266466600161043

95. Chevenet F, Fargette D, Guindon S, Bañuls A-L. EvoLaps: a web interface to visualize continuous phylogeographic reconstructions. BMC Bioinformatics. 2021;22(1):463. https://doi.org/10.1186/s12859-021-04386-z PMID: 34579644

96. Chevenet F, Fargette D, Bastide P, Vitré T, Guindon S. EvoLaps 2: Advanced phylogeographic visualization. Virus Evol. 2023;10(1):vead078. https://doi.org/10.1093/ve/vead078 PMID: 38188280

97. Abubakar Z, Ali F, Pinel A, Traoré O, N'Guessan P, Notteghem J-L, et al. Phylogeography of Rice yellow mottle virus in Africa. J Gen Virol. 2003;84(Pt 3):733–43. https://doi.org/10.1099/vir.0.18759-0 PMID: 12604826