Virus Evolution, 2019, 5(2): vez023

doi: 10.1093/ve/vez023
Research article

Comparing patterns and scales of plant virus phylogeography: *Rice yellow mottle virus* in Madagascar and in continental Africa

Mbolarinosy Rakotomalala,[1,†] Bram Vrancken,[2,†,‡] Agnès Pinel-Galzi,[3] Perle Ramavovololona,[4] Eugénie Hébrard,[3] Jean Stéphan Randrianangaly,[5] Simon Dellicour,[2,6,§] Philippe Lemey,[2,**] and Denis Fargette[3,*]

[1]Centre Régional de Recherche du Nord-Ouest du FOFIFA, BP 289, Mahavoky Avaratra, Mahajanga 401, Madagascar, [2]Department of Microbiology and Immunology, Laboratory for Clinical and Epidemiological Virology, Rega Institute, KU Leuven, Herestraat 49 box 1040, 3000 Leuven, Belgium, [3]IRD, Cirad, Université Montpellier, IPME, 911 avenue Agropolis, BP 64501 34934 Montpellier cedex 5, France, [4]Département de Biologie et d'Ecologie Végétales, Faculté des Sciences, Université d'Antananarivo, BP 906, [5]Direction de la Protection des Végétaux, Ministère de l'Agriculture, BP1042, Antananarivo 101, Madagascar and [6]Spatial Epidemiology Lab, Université Libre de Bruxelles, CP 264 / 3,50 av FD Roosevelt, B-1050 Brussels, Belgium

*Corresponding author: E-mail: denis.fargette@ird.fr

†These authors contributed equally to this work.

‡http://orcid.org/0000-0001-6547-5283

§http://orcid.org/0000-0001-9558-1052

**http://orcid.org/0000-0003-2826-5353

Abstract

Rice yellow mottle virus (RYMV) in Madagascar Island provides an opportunity to study the spread of a plant virus disease after a relatively recent introduction in a large and isolated country with a heterogeneous host landscape ecology. Here, we take advantage of field survey data on the occurrence of RYMV disease throughout Madagascar dating back to the 1970s, and of virus genetic data from ninety-four isolates collected since 1989 in most regions of the country to reconstruct the epidemic history. We find that the Malagasy isolates belong to a unique recombinant strain that most likely entered Madagascar through a long-distance introduction from the most eastern part of mainland Africa. We infer the spread of RYMV as a continuous process using a Bayesian statistical framework. In order to calibrate the time scale in calendar time units in this analysis, we pool the information about the RYMV evolutionary rate from several geographical partitions. Whereas the field surveys and the phylogeographic reconstructions both point to a rapid southward invasion across hundreds of kilometers throughout Madagascar within three to four decades, they differ on the inferred origin location and time of the epidemic. The phylogeographic reconstructions suggest a lineage displacement and unveil a re-invasion of the northern regions that may have remained unnoticed otherwise. Despite ecological differences that could affect the

transmission potential of RYMV in Madagascar and in mainland Africa, we estimate similar invasion and dispersal rates. We could not identify environmental factors that have a relevant impact on the lineage dispersal velocity of RYMV in Madagascar. This study highlights the value and complementarity of (historical) nongenetic and (more contemporaneous) genetic surveillance data for reconstructing the history of spread of plant viruses.

Key words: phylogeography; *Rice yellow mottle virus*; Bayesian statistical framework; molecular clock; dispersal statistics; ecological drivers.

1. Introduction

Less than a decade ago, a methodology for reconstructing the phylogeographic history of viruses on a continuous landscape in a Bayesian statistical framework was introduced (Lemey et al. 2010). It has increasingly been applied to animal viruses (Trovão et al. 2015a; Bourhy et al. 2016; Jacquot et al. 2017; Tian et al. 2018), and in some cases to human viruses (Faria et al. 2012; Pybus et al. 2012; Dellicour et al. 2018a). In contrast, it has been rarely been applied to plant viruses (Monjane et al., 2011; De Bruyn et al., 2012; Trovão et al. 2015b) although plant immobility may result in a regular virus spread, in particular for plant viruses with short-range vectors. Recently however, the spatiotemporal dispersal of *Rice yellow mottle virus* (RYMV) in mainland Africa was inferred using the continuous diffusion model framework (Trovão et al. 2015b). The choice of the spatially continuous model for RYMV spread was motivated by the predominance of short-range transmission (Bakker 1974) and by the high degree of genetic isolation by distance (Abubakar et al. 2003). The spatiotemporal estimates indicated the emergence of RYMV at the end of the nineteenth century in Africa followed by an overall regular dispersal over more than a century across West Africa and across East Africa. A natural extension of studies conducted at the continental scale is to apply the spatially continuous model at a country scale (Levin 1992; Baudelle and Regnauld 2004; Chave 2013). An island sufficiently large for the epidemics to develop and isolated enough to prevent recurrent introductions from blurring the dispersal patterns could provide an interesting model system of study (Cliff and Hagget 1995). For this reason, we focus on RYMV in Madagascar, a 600,000 km^2 large country isolated from mainland Africa by the Mozambique canal that is at least 400 km wide. Rice is the staple crop in Madagascar (Van Chi-Bonnardel 1973; Le Bourdiec 1977) and RYMV is the main disease of rice (Reckhaus and Andriamasintseheno 1997). Scrutinizing the historical spread of RYMV in Madagascar will also help to understand the epidemiology and contribute to control the disease.

RYMV was first isolated in 1966 in Western Kenya near Lake Victoria (Bakker 1974). Since then, RYMV has been detected in almost all rice-producing countries of sub-Saharan Africa (Abo, Sy, and Alegbejo 1997), in Madagascar (Reckhaus and Andriamasintseheno 1997) and recently in the North-West of Ethiopia (Rakotomalala et al. 2014). RYMV currently is a major deterrent to rice cultivation in Africa (Abo, Sy, and Alegbejo 1997) and in Madagascar (Reckhaus and Andriamasintseheno 1997; Rakotomalala et al. 2014) and has not been detected outside this continent. RYMV has a narrow host range restricted to the two cultivated rice species *Oryza sativa* and *O.glaberrima*, the wild rice species *O.longistaminata* and *O.barthii* and a few related wild *Poaceae* species (Bakker 1974). The virus is transmitted by chrysomelid beetles (Bakker 1974), by mammals (Sarra and Peters 2003), and by contact during cultural practices (Traoré et al. 2006). No evidence of seed transmission has been found in cultivated and wild rice species (Bakker 1974; Fauquet and Thouvenel 1977; Konaté, Sarra, and Traoré 2001; Allarangaye et al. 2006).

RYMV is a single-stranded positive-sense RNA species of the *Sobemovirus* genus in the *Solemoviridae* family (King et al. 2018) with a ca. 4,450-nucleotide-long genome organized into five open reading frames (ORFs) (Sõmera, Sarmiento, and Truve 2015). ORF1, located at the 5′ end of the genome, encodes a small protein (P1) involved in virus movement and in gene silencing suppression. ORF2, which encodes the central polyprotein, has two overlapping ORFs. ORF2a encodes a serine protease and a viral genome-linked protein that determines the virulence. ORF2b, which is translated through a −1 ribosomal frameshift mechanism as a fusion protein, encodes an RNA-dependent RNA polymerase. ORF4 is translated from a subgenomic RNA at the 3′ end of the genome and encodes the coat protein (CP). Recently, the presence of a fifth ORF (ORFx), conserved in all sobemoviruses, which overlaps the 5′ end of the ORF2a in the +2 reading frame, was reported (Ling et al. 2013). The CP gene (ORF4), 720 or 723 nucleotide long (depending on the strain), has been used for the phylogeographic reconstructions. It revealed a high (up to 20% between isolates) and spatially structured diversity with distinct strains in East and West Africa (Abubakar et al. 2003). Thus far, there is no evidence of intra-ORF4 recombination, with the exception of a localized strain recently found in Western Kenya (Adego et al. 2018). RYMV isolates of Madagascar belong to a specific strain named the S4-Mg strain (Rakotomalala et al. 2013; Ochola et al. 2015).

Due to its economic importance, rice in Madagascar has been subjected to regular phytosanitary surveys, more frequent and comprehensive than in most countries of continental Africa. The first reports of RYMV symptoms by farmers and agricultural advisers date back to the 1970s. In several instances, leaf samples with symptoms were kept dried for years and were found positive a decade later when the first serological tests became available (S. Randrianangaly, unpublished data). By the 1980s, plant pathologists conducted regular surveys throughout the country. Since the 1990s, RYMV presence was not only assessed through symptom expression, but also after mechanical inoculation of susceptible cultivars or by serological agar-gel double-diffusion tests (Reckhaus and Randrianangaly 1990). We resumed these surveys between 2003 and 2017 in which RYMV was diagnosed by ELISA and molecular tests. Serological or molecular tests have shown that RYMV is not detected in symptom-free plants, even those collected close to or alongside infected plants. Together, these surveys provide reliable information on the presence/absence of the virus in the visited fields. Results of the field surveys are accessible in scientific journals (Reckhaus and Randrianangaly 1990; Reckhaus and Andriamasintseheno 1997) or in internal reports.

Analyses of ORF4 sequences from isolates collected since 1966 revealed that RYMV is a fast-evolving pathogen (Fargette et al. 2008). This implies that samples collected over the last few

decades may contain information about recent evolutionary dynamics. Such temporally spaced sequence data can, however, only reliably inform evolutionary rate estimates if genetic divergence over the sampling time range, or the so-called temporal signal, can be identified and statistically supported (Drummond et al. 2003), and it is now well-recognized that this assumption should be thoroughly tested in order to avoid spurious results (Firth et al. 2010; Duchêne et al. 2015; Murray et al. 2016; Rambaut et al. 2016). How accurately the spatial patterns of spread can be recovered by phylogeographic methods, on the other hand, remains challenging to evaluate. Recently, key aspects of the phylogeography of RYMV have tentatively been linked to the spatially heterogeneous and rapidly changing rice cultivation in mainland Africa (Pinel-Galzi et al. 2015; Trovão et al. 2015b), a historically (Portères 1950)—and now genetically (Ly et al. 2018; Choi et al. 2019)—well documented crop in Africa. However, this link was not significantly supported when appropriate negative controls were used (Dellicour et al. 2018b).

To provide information on the genetic and geographical origin of RYMV in Madagascar, and on whether or not the virus introduction was unique, we examined the phylogenetic relationships of the S4-Mg strain in the light of an expanded full-length sequence dataset that included new strains from East Africa (Rakotomalala et al. 2014; Ndikumana et al. 2017; Adego et al. 2018). By capitalizing on available and newly generated ORF4 sequence data, we also investigated the patterns and dynamics of the spatiotemporal spread of RYMV in Madagascar, we contrasted the phylogeographic inferences against historical knowledge of the RYMV spread in Madagascar, and analyzed the determinants of the velocity with which RYMV invades new regions and migrates within infected areas.

2. Materials and methods

2.1 Surveys and sampling

Madagascar is the fifth largest island in the world. It covers 600,000 km^2, stretches 1,500 km from south to north, 500 km from east to west, and is isolated from mainland Africa by the Mozambique canal that is at least 400 km wide. The landscape is divided into three zones: a high-altitude plateau (1,000–2,500 m) forms a south-to-north barrier between the wide plains at the west and a narrow coastal zone at the east. The island has several climatic zones, tropical humid at the east, tropical savanna at the north-west, warm semi-arid at the south and south-west, and a subtropical oceanic highland climate in the high plateau of the center of the country (Donque 1975; Peel, Finlayson, and McMahon 2007). With 1,500,000 ha of rice cultivation, rice is the main staple crop of Madagascar. It occupies nearly half of the total cultivated area. Rice is grown in six zones: the north, north-west, and central-western regions, the central part of the Malagasy highlands, the east, and the central-eastern part including the Lake Alaotra basin with its swampy areas, plains, and valleys suited for rice. Rice is also cultivated to a smaller extent in the south of the country. There are four main rice cultivation modes in Madagascar: irrigated rice, rainfed lowland rice, rainfed upland rice, and rice as a first crop after slash and burn. In ratio of cultivated area, irrigated rice is the most important, covering 82 per cent of all area under rice in 2008. About 60 per cent of irrigated rice is transplanted.

A total of eighty-eight isolates were collected between 2003 and 2017 throughout Madagascar, except from the difficult-to-access Melaky region in the west of the country. The outbreaks were sporadic in the high-altitude plateau in the center of the country. In the extreme south of the country, the disease was detected in 2010 only. We also had access to six isolates collected during surveys in 1989 and 1991 from the Lake Alaotra basin, the Marovoay plains, and the high-altitude plateau near Antananarivo.

2.2 ORF4 and full-length genome RYMV sequence datasets

ORF4 sequences (720 nt) of eighty isolates collected between 1989 and 2010 in Madagascar were retrieved from NCBI GenBank (Rakotomalala et al. 2013). Additionally, ORF4 sequences of fourteen Malagasy isolates, collected in 2014 and in 2017, were newly generated as described previously (Pinel et al. 2000) and deposited at NCBI GenBank under the accession numbers MK94093–MK94106. All Malagasy isolates were annotated with the date (year) and location (latitude and longitude) of sampling. The ninety-four aligned ORF4 sequences are referred to as the Mg94 dataset.

The ORF4 sequence dataset of isolates from mainland Africa was assembled by retrieving the publicly available sequences from NCBI GenBank (28 December 2018). The dataset consists of 450 geo-referenced sequences (720 nt) collected between 1966 and 2017 in twenty-two countries from West Africa (eleven countries), Central Africa (six countries), and East Africa (five countries), that is from almost all rice-growing countries of sub-Saharan Africa. The sequences of the isolates from countries at the west of Central Africa (Cameroun, Central African Republic, Chad) were analyzed together with those from West Africa (Benin, Burkina Faso, Ghana, Guinea, Ivory Coast, Mali, Niger, Nigeria, Sierra-Leone, Togo) and constitute the West-African dataset. The sequences of the isolates from countries at the east of Central Africa (east of the Republic Democratic of the Congo, Burundi, Rwanda) were analyzed with those from East Africa (Ethiopia, Kenya, Malawi, Tanzania, Uganda) and make up the East African dataset. The West African and the East African datasets were analyzed separately because of the independent epidemiological dynamics in the two regions (Trovão et al. 2015b). Compared to the 300 taxon dataset analyzed earlier by Trovão et al. (2015b), the updated dataset of 450 sequences included 30 additional isolates from West Africa (210 vs. 180), referred to as the WA210 dataset, and twice as many isolates from East Africa (240 vs. 120) referred to as the EA240 dataset. In West Africa, the newly sequenced isolates belong to recorded strains. In contrast, the extension of the surveys in East Africa revealed new strains at the north of Ethiopia (Rakotomalala et al. 2014), at the east of the Victoria Lake basin (Uke et al. 2015; Adego et al. 2018), and at the south of Malawi (Ndikumana et al. 2017).

The full-length genome dataset was assembled by retrieving the publicly available sequences from NCBI GenBank (28 December 2018). The dataset consists of thirty sequences, twenty-four from East Africa, and six from Madagascar. Compared to the twenty-two taxon full-length dataset used in earlier studies (Ochola et al. 2015), this updated dataset included two additional full-length sequences of the S4-Ug strain found in western Kenya (Adego et al. 2018), one sequence from the S4-Lv strain (Adego et al. 2018), and five sequences of isolates of three recently identified strains: two sequences of the S4-Et strain from Ethiopia (Rakotomalala et al. 2014), two of the S4-Ke strain from Kenya (Adego et al. 2018), and one S7 sequence from Malawi (Ndikumana et al. 2017).

2.3 Genetic, phylogenetic, and recombination analyses

Sequences were aligned using CLUSTAL X with default parameters (Thompson, Higgins, and Gibson 1994). The diversity index (Pi), which is the average number of nucleotide substitutions per site between any two sequences in a multiple sequence alignment, was calculated for the Malagasy and continental Africa ORF4 (720 nt) sequence datasets using DNASp version 6 (Rozas et al. 2017). Maximum-likelihood (ML) phylogenetic trees were inferred using the PHYML-3.1 algorithm implemented in the SEAVIEW version 4.7 software (Gouy, Guindon, and Gascuel 2010) under a HKY85 substitution model. The ML trees were rooted at the point in the tree that minimizes the variance of the root-to-tip distances. The approximate likelihood ratio test (aLRT) was applied to assess the reliability of key branches of the trees.

The full-length sequence alignment was screened for recombination signals using the RDP version 4.95 software (Martin et al. 2015). 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 0.001. Only recombination events detected by more than three of the seven methods implemented in RDP were considered. In addition, the genetic distance across the genome between the S4-Mg strain, the other S4 strains, and the S5 sister-strain, all originating from East Africa, were visualized by SIMPLOT (Lole et al. 1999).

2.4 Tests of temporal signal

A suite of tests was applied to assess the significance of the temporal signal in the sequence datasets (Duchêne et al. 2015; Murray et al. 2016). A ML phylogenetic tree was reconstructed under a HKY85 model with a Γ-distributed rate variation among sites as implemented in SEAVIEW (Gouy, Guindon, and Gascuel 2010). We used TempEst v1.5.1 (Rambaut et al. 2016) to regress phylogenetic root-to-tip distances against sampling date using the root that minimized the residual mean squares. The significance of the regression was assessed by random permutation of the sampling dates over the sequences (tip permutations), using the correlation coefficient as the test statistic. A total of 1,000 replicates of the data with the sampling dates randomly permuted were generated to obtain the expectation of the test statistic in the absence of temporal signal. The *P*-value is the proportion of replicates with a test statistic greater than or equal to the test statistic obtained using the observed sampling dates. The significance of the temporal signal was also evaluated while accounting for the nonindependence of the data. To this end, the mean rate and its 95 per cent highest density probability (HPD) estimated with the observed sampled dates using BEAST v1.8.4 were compared to a null distribution obtained by randomly permutating the tip dates ten times (Firth et al. 2010). The temporal signal analyses were also conducted when randomizing the sampling dates over clusters of taxa (cluster permutations) and not over individual taxa (tip permutations), with clusters defined as monophyletic clades for which all taxa were sampled in the same year (Duchêne et al. 2015; Murray et al. 2016).

2.5 Bayesian evolutionary inference

We reconstructed time-calibrated epidemic histories using a Bayesian statistical framework implemented in the software package BEAST version 1.8.4 (Drummond et al. 2012). BEAST uses Markov chain Monte Carlo (MCMC) integration to average over all plausible evolutionary histories for the data, as reflected by the posterior probability. All analyses were performed using the Broad-platform Evolutionary Analysis General Likelihood Evaluator (BEAGLE) library to enhance computation speed (Suchard and Rambaut 2009; Ayres et al. 2012).

To model the nucleotide substitution process, the codon positions were partitioned into first + second vs. third positions. A separate HKY85 substitution model was applied, with a discretized Γ distribution (HKY + Γ) to model rate heterogeneity across sites to each of the two codon partitions (Shapiro, Rambaut, and Drummond 2006). To accommodate among-lineage rate variation, an uncorrelated relaxed molecular clock that models the branch rate variation according to a lognormal distribution was specified (Drummond et al. 2006). The flexible nonparametric demographic skygrid prior was selected (Gill et al. 2013). Stationarity and mixing (e.g. based on effective sample sizes >200 for the continuous parameters) were examined using Tracer version 1.6. The MCC trees were summarized using TreeAnnotator.

To study the geographic spread of RYMV in continuous space and to quantify its tempo and dispersal, we fitted separate continuous phylogenetic diffusion models to the Madagascar, East and West African datasets, modeling the change in coordinates (latitude and longitude) along each branch in the evolutionary history as a bivariate normal random deviate (Lemey et al. 2010). As an alternative to homogeneous Brownian motion, we adopted a relaxed random walk (RRW) extension that models branch-specific variation in dispersal rates by independently drawing branch-specific rate scalers of the RRW precision matrix from a Cauchy distribution to relax the assumption of a constant precision (= 1/variance) among branches (Lemey et al. 2010). Bayesian inference using continuous diffusion models yield a posterior distribution of phylogenetic trees, each having ancestral nodes annotated with location estimates. We summarized several dispersal statistics from the posterior estimates of the continuous phylogenetic diffusion process (Pybus et al. 2012). Specifically, we provided mean posterior estimates and the 95 per cent HPD intervals for the weighted branch dispersal velocity (km/year) (Dellicour et al. 2016), which quantifies how fast a virus migrates within the area it affects. The dispersal statistics were summarized across successive time periods. In addition to dispersal rates, we also visualized the epidemic wavefront as the largest great-circle distance traveled from the root location.

2.6 Impact of environmental factors on lineage dispersal velocity

We analyzed the impact of several environmental factors on the lineage dispersal velocity of RYMV in Madagascar using R functions available in the package 'seraphim' (Dellicour et al. 2016). In this analysis, each environmental factor was described by a raster that defines its spatial heterogeneity and that was used to compute an environmental distance for each branch in the phylogeny using two different path models: (1) the least-cost path model, which uses a least-cost algorithm to determine the route taken between the starting and ending points (Dijkstra 1959), and (2) the Circuitscape path model, which uses circuit theory to accommodate uncertainty in the route taken (McRae 2006; McRae et al. 2008). Here, we investigated the impact of rice area harvested, elevation, mean annual temperature, and annual precipitation (see Supplementary Table S2 for the sources of the original raster files). These rasters were either tested as potential conductance factors (i.e. factors facilitating movement) and/or as potential resistance factors (i.e. factors

impeding movement). Correlations between phylogenetic branch durations and environmental distances were estimated with the statistic Q defined as the difference between two coefficients of determination (R^2): (1) R^2 obtained when branch durations are regressed against environmental distances computed on the environmental raster, and (2) R^2 obtained when branch durations are regressed against environmental distances computed on a null raster, that is an environmental raster with a value of '1' assigned to all the cells. Statistical support was then evaluated against a null distribution generated by a randomization procedure and formalized as a Bayes factor (BF) value (Dellicour et al. 2017). To account for the uncertainty related to the Bayesian inference, this analysis was based on 500 trees sampled from the post-burn-in posterior distribution inferred using the continuous phylogeographic model.

3. Results

3.1 Genetic characteristics of the RYMV strain from Madagascar

We first identified where in the genome recombination occurred, and which were the likely strains involved, by plotting the genetic distance between sequences as a function of the position in the genome. Such similarity plots, implemented in SIMPLOT, were made for the full genome of the six Malagasy isolates with representative isolates of the other S4 strains and of the S5 sister-strain. As the results were consistent among the six isolates from Madagascar, they are illustrated for only a single isolate (isolate Mg15) (Fig. 1). This identifies three putative recombination events that were also detected by at least five of the seven recombinant detection methods implemented in RDP. A ML phylogenetic tree from each of the four fragments was reconstructed using PHYML (Fig. 2). In three of the four recombinant regions (regions II, III, and IV), the S4-Mg isolates are most closely related to the S4-Ug strain (Fig. 2), and in fragment IV—which includes the ORF4—the S4-Mg isolates are nested within the S4-Ug strain. In contrast, in region I the S4-Mg isolates are most closely related to the S4-Et strain.

The six sequences of RYMV Madagascar belonged to a monophyletic clade with a high branch support whatever the genome segment considered (Fig. 2). The virus diversity estimated from the ORF4 sequences of ninety-four isolates from Madagascar (600,000 km^2) was lower than that in other rice producing countries of East and West Africa estimated from a comparable subset of sequences sampled at the a similar spatial scale. The average pairwise distance is 1.6 per cent for the Madagascar vs. 3.3 per cent in Ivory Coast (fifty-one isolates; 320,000 km^2), 4.3 per cent in Mali (forty-one isolates; 1,240,000 km^2, but rice is cultivated in the southern half of the country only), and 8.3 per cent in Tanzania (154 isolates; 950,000 km^2).

3.2 Temporal signal

Regression analysis of root-to-tip distances for the ninety-four ORF4 sequences from Madagascar (Supplementary Fig. S1) indicated that the sampling dates were a significant predictor ($P = 0.02$). In Bayesian dating analyses, there was no overlap between the 95 per cent HPD intervals of the substitution rate estimated from the real sampling dates and those estimated after tip permutations (Fig. 3). Using clustered permutations (thirty-eight clusters), however, reduced the *P*-value of the linear regression to 0.05. Moreover, the mean substitution rate estimated from the real sampling dates in Bayesian dating analyses fell within the range of values obtained after cluster permutations (Fig. 3). Hence, the temporal signal in the Madagascar dataset (ninety-four sequences) may not be sufficient to reliably estimate the substitution rate of RYMV in Madagascar.

For the full dataset (544 sequences, including the ninety-four sequences from Madagascar), the linear regression between the sampling dates and the root-to-tip distances of the 544 isolates was significant after tip ($P < 10^{-3}$) and cluster permutations (340 clusters, $P = 5.\ 10^{-3}$). The estimated mean rate was 9.2×10^{-4} substitution per site per year (Supplementary Fig. S1). In

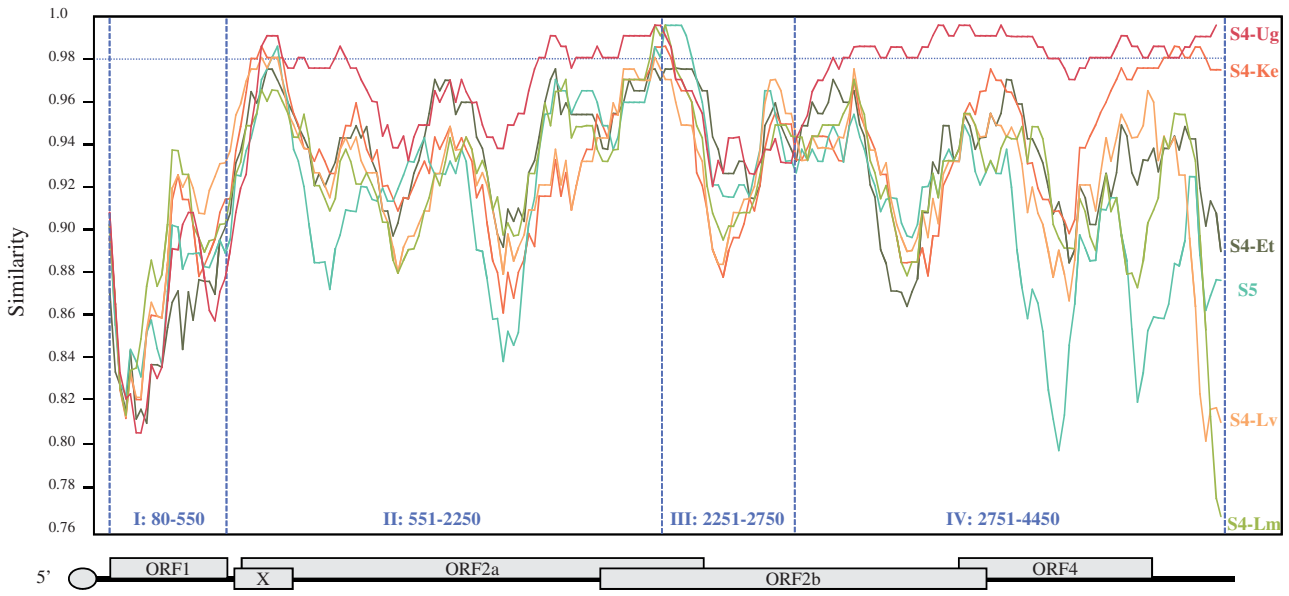

Figure 1. Similarity plot of the full-length sequence of a S4-Mg isolate (Mg15) with East African isolates representative of the other S4 strains: S4-Lv (Tz5), S4-Lm (Tz8), S4-Ug (Ug207), S4-Et (Et5), S4-Ke (Ke101), and of the S5 sister-strain (Tz3). Genetic distances were estimated using the Kimura two-parameter model in a sliding window of 200 nucleotides and a step size of 20 (top). The three putative recombination sites are identified where sequence crossover occurs and are indicated by vertical dotted lines. The horizontal dotted line is set at the 0.98 identity level to indicate the S4-Mg strain diversity. The genomic organization of RYMV is sketched below the similarity plot.

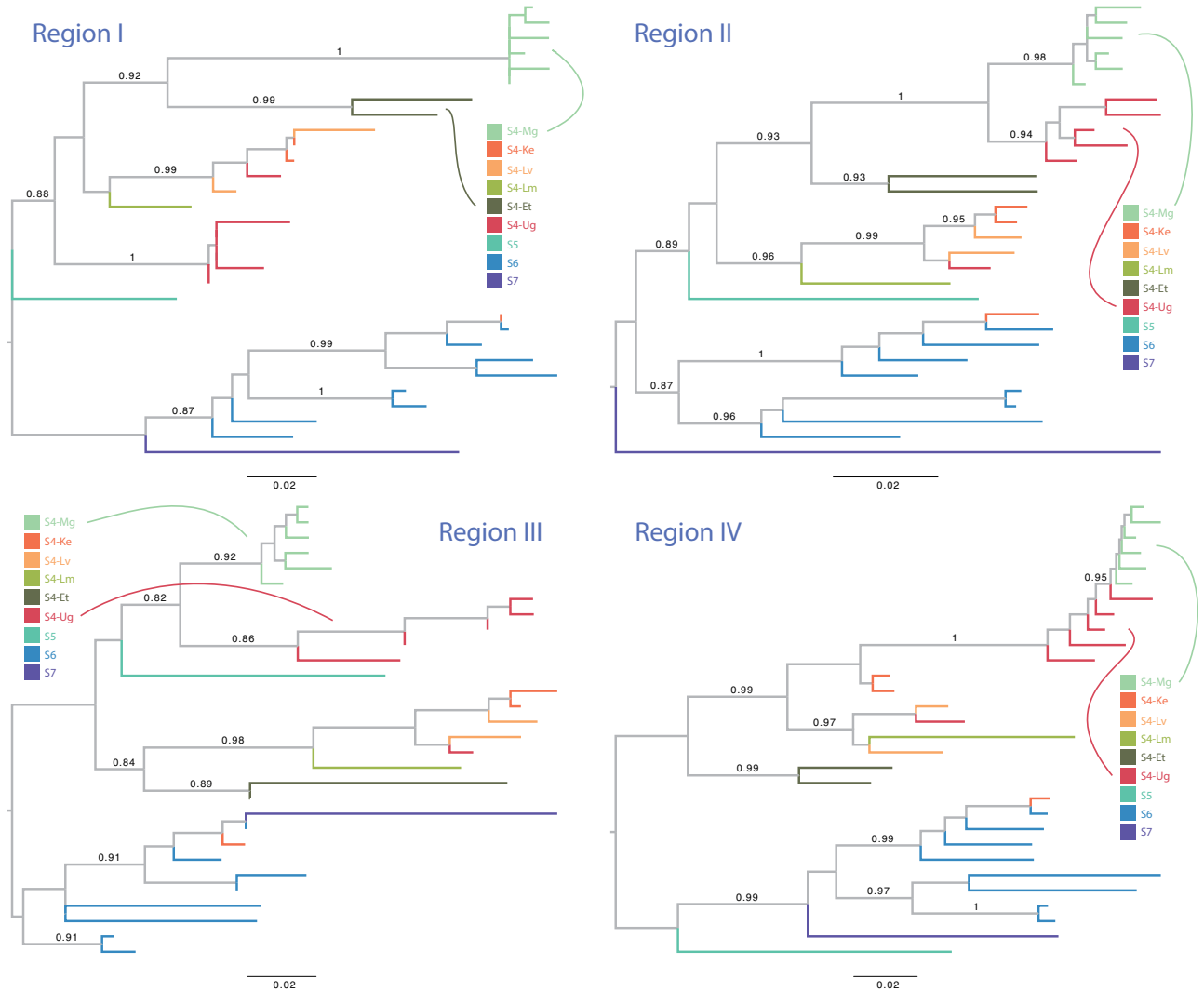

Figure 2. Phylogenies estimated from genomic fragments I to IV (see Fig. 1) of the thirty complete sequences of isolates representative of the RYMV strains of East Africa and Madagascar. The trees were midpoint rooted, and numbers next to branches indicate their support based on the aLRT test. Tips are colored by strain, with color-strain correspondence as in the legend. The scale bar is in units of substitutions per site.

Bayesian dating analyses, the mean estimated substitution rate was 1.1×10^{-3} substitution per site per year with a 95 per cent HPD of 8.8×10^{-4}–1.3×10^{-3} that did not overlap with those obtained either after tip or after cluster permutations (Fig. 3). These results indicate that a reliable estimate of the evolutionary rate can be obtained by combining the temporal signal of the East Africa (EA240), West Africa (WA210), and Malagasy (Mg94) datasets. To achieve this while independently modeling the phylogeographic diffusion process in each area, the East Africa, West Africa, and Madagascar datasets were considered as separate partitions with independent tree topologies and evolutionary parameters, except for the mean clock rate that was shared among datasets. This model was applied to reconstruct the phylogeography of RYMV and to estimate the dispersal statistics in the three geographical partitions.

3.3 Field surveys and phylogeographic reconstructions

In the early 1970s, in the very north of Madagascar (Antsiranana, Ambanja, Ambilobe, Antsonihy), vernacular names (such as Menamiretaka) were given by farmers to rice plants showing RYMV-like symptoms (Supplementary Table S1 and Fig. 4). In the first half of the 1980s, disease symptoms were observed in the large rice plains in the northwest of the country (Betsiboka valley, Marovoay, Ambato-Boeny, Mampikony). It was later confirmed by serological tests that the symptoms resulted from RYMV infection (see Section 2). In 1990, RYMV was detected by biological and serological tests in symptom-carrying plants from this region. In the second half of the 1980s, symptoms were observed in the northeast of Madagascar (Vohemar, Sambava, Andapa) and in the central west (Miandrivazo). In 1989, the virus was detected in the Lake Alaotra basin, in the central east, the largest rice producing region of Madagascar (100,000 ha). In the first half of the 1990s, symptoms were observed with subsequent virus detection in the central west (Morondava, Maintirano, Melaky district), in the high plateau (Antananarivo), and along the east coast (Vavatenina plains, Toamasina, Mananjary). In the second half of the 1990s, the disease was observed further southeast (Tsivory). Since 2000, the disease has been observed further south, in the southwest (Tandava) and in the southeast (Manakara). RYMV in Madagascar is now present in regions

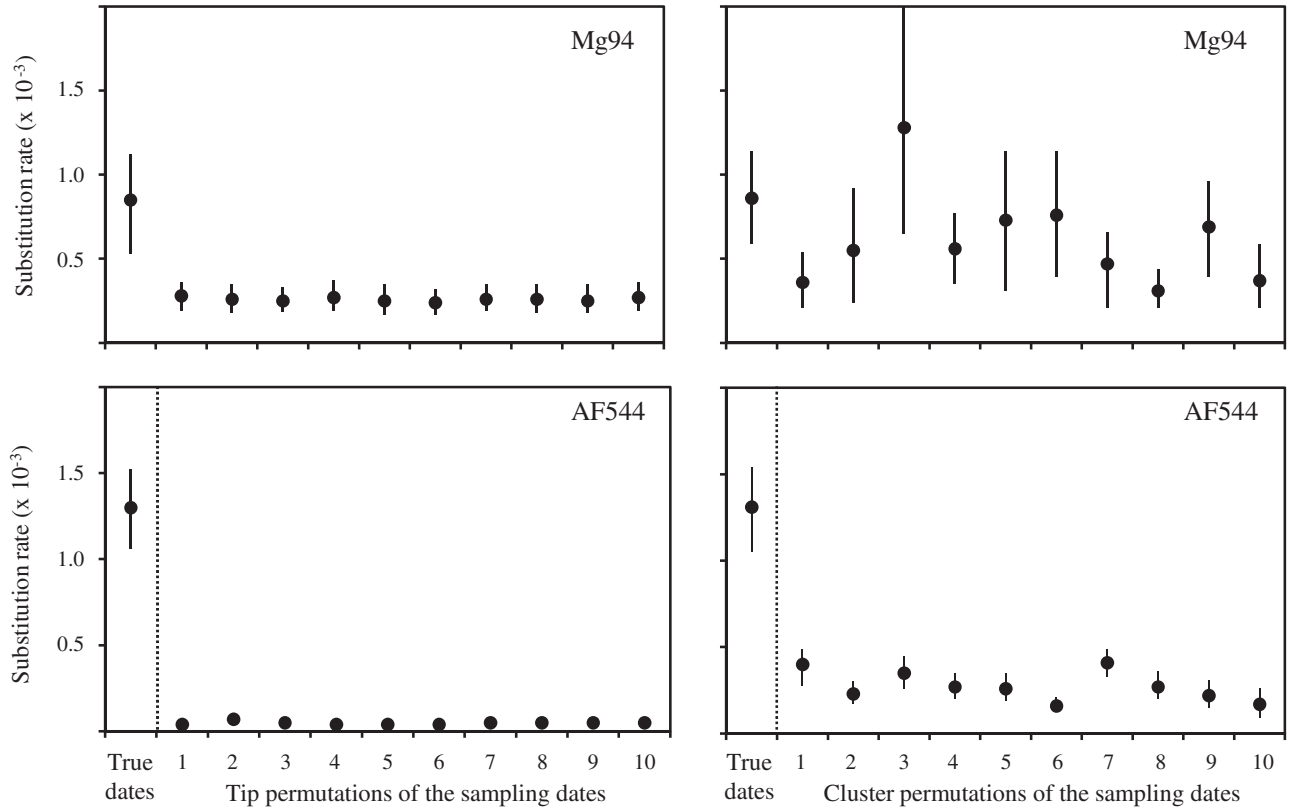

Figure 3. Estimates of the mean and of the 95 per cent HPD interval of the substitution rate (subs/site/year) in the AF544 and the Mg94 datasets from the sampled dates, and from the dates after tip or after cluster permutations.

with different climates and where rice is cultivated under different modes (see Section 2), except in the very south of the country (Tsiombe, Beloha, Ampanihy, Ejeda) where it is absent although surrounded by locations where the disease has been present for years. In the high-altitude plateau, the disease remained sporadic, observed near Antananarivo in 2014 only since its first report in 1991. Furthermore, the disease was not found in locations neighboring Antananarivo during surveys conducted the same year (Supplementary Table S1).

This sequence of observations (Fig. 4), the low genetic diversity of the Malagasy isolates and the well-supported monophyly of the S4-Mg strain (Fig. 2) are consistent with a single introduction of RYMV in the early 1970s in the very north of Madagascar, followed by a gradual spread southward along the east and along the west parts of the country. Accordingly, the disease has crossed ca. 1,500 km from the north to the south of Madagascar within ca. 40 years, so with an average of about 37 km/year.

The reconstructed epidemic history based on the virus genetic data places the time to the most recent common ancestor (t_{MRCA}) of the sampled RYMV strains in Madagascar around 1983 (95% HPD: 1979–87) with an initial spread from the northwestern and Lake Alaotra basin to the other regions of the country (Fig. 4). Of note, this date range largely overlaps with the t_{MRCA} estimate derived from the Madagascar dataset alone (1981, 95%CI: 1975–86). The southwards dispersal occurred along the west and east sides of the country, with few virus exchanges across the high-altitude plateau between the east and west regions. In the high-altitude plateau, where the disease is sporadic, closely related isolates were found nearby the capital Antananarivo at the early (1991) and late stages (2014) of RYMV epidemics. This suggests localized low-level virus persistence in the high plateau rather than recurrent and transient introductions from the Lake Alaotra basin, 200 km north, through rice corridors connecting the two regions. Whereas the phylogeographic reconstructions do not infer an origin of the epidemic in the very north of the country, they infer a re-invasion of the northeast around 1995 and of the northernmost part of the country around 2005. The full extent of spread throughout the country was achieved within three decades. The estimated mean wavefront rate was 33 km/year (95% HPD: 24.6–41.3) which is close to that of 37 km per year we deduced from the field surveys.

In Table 1, we compare the weighted average dispersal velocity of RYMV in mainland Africa and Madagascar epidemics. Over the time frame of the Madagascar epidemic (the 1983–2017 period, i.e. a 35-year-long time slice), the lineage dispersal velocity is similar in Madagascar and East Africa (ca. 14–15 km/year) and somewhat lower in West Africa (ca. 11 km/year) although the uncertainty on the estimates overlaps substantially. At earlier 35-year-long time slices, the lineage dispersal velocity in Madagascar was higher than that of the East Africa epidemic. The mean estimates show a similar trend for West Africa, but the large uncertainty that overlaps with that for the Madagascar estimate prevents us from drawing any firm conclusions about these differences. A plot of the epidemic wavefront over time shows that the slope is comparable among the three epidemics in the three decades before the wavefront reaches its plateau (Fig. 5).

The analyses of the potential impact of environmental factors on viral lineage dispersal velocity did not reveal any notable footprint (Supplementary Table S2). Only the elevation raster

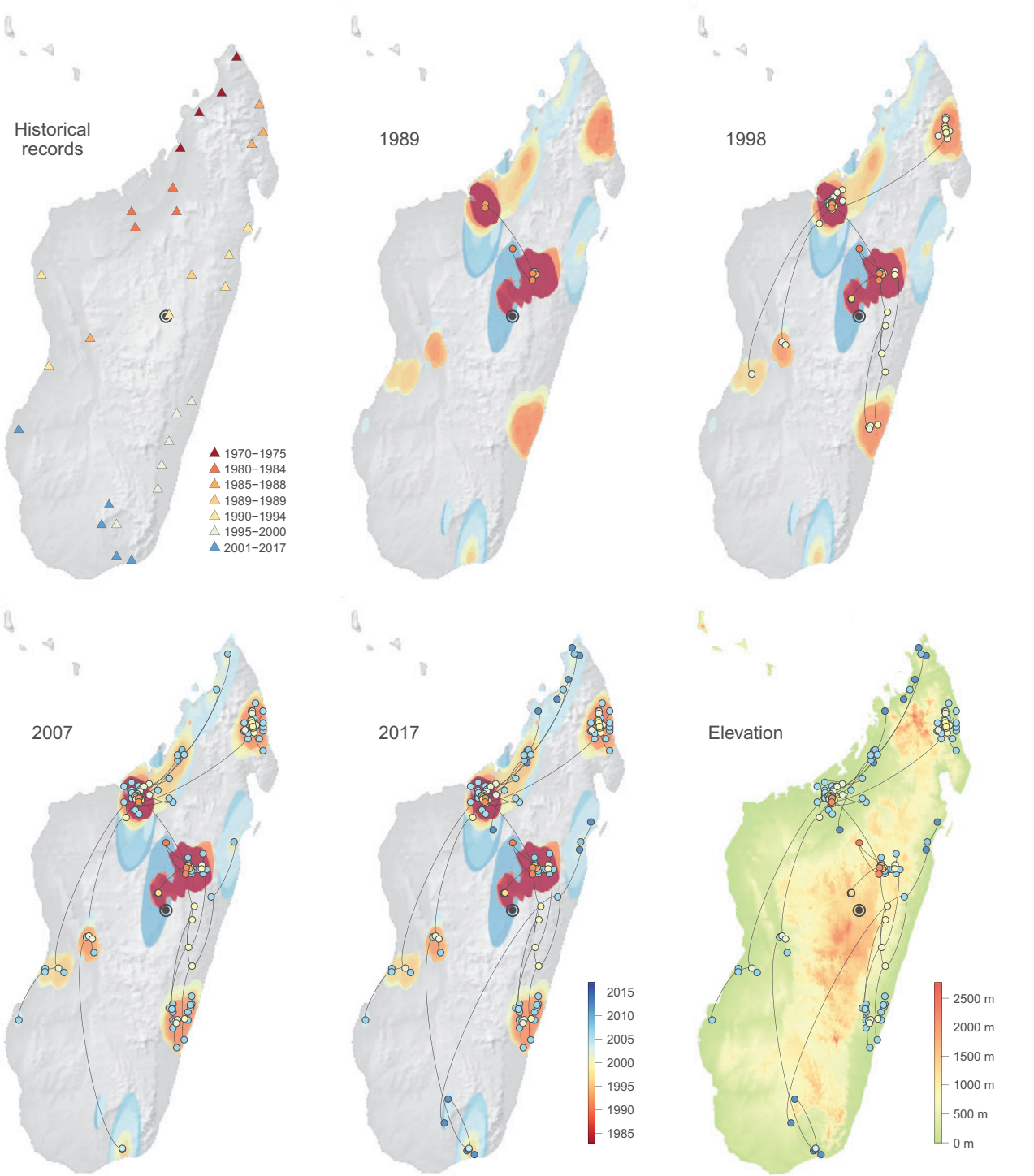

Figure 4. Reconstruction of the continuous spatiotemporal dispersal of RYMV in Madagascar shown from 1989 to 2014 at intervals that capture the major dispersal events. Black lines show a spatial projection of the representative phylogeny. Colored clouds represent statistical uncertainty in the estimated locations of RYMV interval nodes (95% HPD intervals). The first map displays the historical records, that is date and location of the first reports of RYMV symptoms in the different regions of Madagascar; see text and Supplementary Table S1 for details.

tested as a conductance factor with the least-cost path model led to both a positive distribution of Q values and an associated BF support >3 (BF = 3.8). This means that under the least-cost path model, phylogenetic branch durations tend to be more correlated with environmental distances computed on the elevation raster (treated as a conductance raster) than with environmental distances computed on the null raster. However, while the Q distribution is supported, the mean Q value remains rather low. A low Q value indicates that the tested environmental layer (here the elevation layer) does not better explain the heterogeneity in lineages dispersal velocity than a uniform layer (null raster) that only accounts for the impact of

Table 1. Weighted branch dispersal velocity (mean value and 95% HPD).

	Madagascar	West Africa	East Africa
Time slices[a]			
0–35	14.7 [10.6–19.2]	10.5 [8.5–12.5]	13.6 [11.2–16.2]
35–70	–	12.1 [7.5–16.6]	8.4 [5.9–10.7]
70–105	–	9.1 [1.4–17.4]	4.4 [1.4–7.9]
105–140	–	5.7 [0–21.2]	3.7 [0.1–10.2]

Values in brackets represent 95 per cent HPD interval.
[a]Time slices in number of years before 2015.

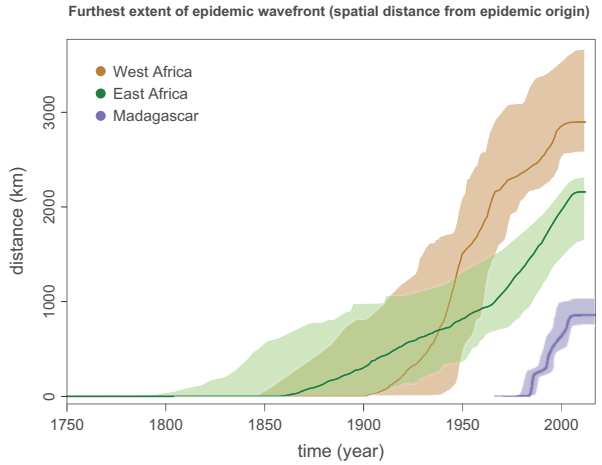

Figure 5. Mean wavefront distances (spatial distances from epidemic origin) in West Africa, East Africa and Madagascar. Mean values are indicated by dark lines and the 95 per cent intervals by colored shadows.

geographic distance. Indeed, using the elevation raster to compute environmental distances with the least-cost path model only increases the correlation with branch durations by ~2 per cent, relative to the null raster. Therefore, if high altitude was directly or indirectly associated with higher lineage dispersal velocity, it contributed only marginally to the observed heterogeneity in RYMV dispersal velocity.

4. Discussion

The geographical distribution of RYMV in continental Africa (Pinel-Galzi et al. 2015) shows that RYMV can persist in diverse climates, ranging from equatorial, monsoon, tropical savanna, humid subtropical climates to warm semi-arid climates (Peel, Finlayson, and McMahon 2007). As the spatial distribution of strains in West Africa is structured according to the main climatic zones, it had been hypothesized that this reflected strain adaptation to specific climatic/ecological conditions. Accordingly, these strains had sometimes been referred to as Sahelian strain, savanna or forest strains (N'Guessan et al. 2001). The RYMV epidemiology in Madagascar does not support this hypothesis. Rather, it shows that a single strain, with a restricted genetic diversity, can rapidly spread and infect rice in diverse climates similar to those in continental Africa. Consequently, the spatial structure of the strains in West Africa more likely reflects the patterns generated by successive founder effects and neutral epidemiological processes (Pinel-Galzi et al. 2015; Trovão et al. 2015b) rather that adaptation to climates. Although wide, the ecological plasticity of the disease in Madagascar is less than that of rice, its main host species. In particular, RYMV was sporadic in the high-altitude plateau (subtropical oceanic highland climate) and absent in the extreme south of the country with a warm semi-arid climate. In these two regions, the susceptibility of the rice cultivars and the presence of vector beetle species (Rakotomalala et al. 2008) suggest that environmental conditions may be unfavorable to disease spread. For instance, low temperature in the high plateau and drought in the south may hinder vector movements and disease spread.

The S4-Mg strain is a combination of genetic fragments of lineages ancestral to the S4-Ug strains (CP, protease, VPg, and polymerase genes) and an ancestral S4-Et strain (P1 gene), assembled through consecutive recombination events. The S4-Ug and the S4-Et strains are found, respectively, at the East of the Lake Victoria basin and in the northwest of Ethiopia. In contrast, the S4-Mg strain is genetically quite distant from the strains S4-Lm, S5, S6, and S7 found along the Tanzanian coasts, and RYMV has not yet been observed in Mozambique (Séré and Ndikumana, pers. comm.). Taken together, this indicates that RYMV in Madagascar was not introduced from the closest continental coasts 400–1,000 km away. Rather, the available data suggest that RYMV virus in Madagascar originates from the very East of Africa. The long distance separating Madagascar from mainland Africa and in particular from the northwest of Ethiopia (>1,000 km) points to a mode of transmission not involving the usual short-range beetle-driven dissemination. This long-range, likely unique, introduction, suggests a human-mediated mode of transmission (Jones 2018), possibly related to rice trade between continental Africa and Madagascar (Le Bourdiec 1977), although any association remains speculative in the absence of a clear pattern.

RYMV in Madagascar presents an opportunity to study the spread of a plant virus after a single and relatively recent introduction in a large country with a heterogeneous host ecology and landscape. The field surveys and the phylogeographic reconstructions converge on a picture of a regular and rapid southward invasion of the epidemic throughout Madagascar within three to four decades. The most notable difference between the reconstructions of the epidemic lies in the early stages as the presence of the disease in the 1970s in the very north of Madagascar reported by farmers is not captured by our sample of the epidemic. As our sequence dataset contains only a limited number of isolates from this region, it remains difficult to distinguish between the scenario where the initial isolates from the north of Madagascar have no surviving progeny, or that their descendants were not sampled. In addition, the phylogeographic reconstructions capture a re-invasion of the north regions that is difficult to distinguish from ongoing local transmission by traditional surveillance methods. The fast dissemination of RYMV in Madagascar underlines the risk of a rapid diffusion of virulent pathotypes. For instance, isolates which readily overcome the *Rymv1-2* allele of high resistance were found in 2003 in the north-west of Madagascar (Rakotomalala et al. 2008). With a dispersal of around 15 km/year, these virulent isolates can rapidly become, everywhere in the country, a threat to rice cultivation with varieties harboring this resistance genotype. Such risks should be considered in the design of RYMV control strategies (Hébrard et al. 2018).

The spread of RYMV in Madagascar was faster than that in mainland East Africa at the early stages of the epidemics when the mode and the extent of rice cultivation were unfavorable to virus spread. The dispersal rate increased over time in

mainland Africa, possibly reflecting the intensification of rice cultivation favorable to virus spread (Abo, Sy, and Alegbejo 1997), so the difference in dispersal velocities between mainland Africa and Madagascar fades towards the present. This is still unexpected considering that several factors can favor RYMV dispersal in Madagascar even more: the dense and continuous rice landscape, endemic beetle vector species such as *Dicladispa gestroi* (Delucchi 2001), the regular use of the zebu which disseminates the virus during trampling and through feces, and the high rate of transplanted rice (Reckhaus and Andriamasintseheno 1997). Yet, these differences do not appear to affect the rate of RYMV spread into new areas as the slope of the epidemic wavefront distance over time is comparable among the three epidemics in the three decades before the wavefront reaches its plateau (Fig. 5). Combined, this suggested the existence of a threshold in rice density, currently reached both in Madagascar and in continental Africa, beyond which the virus dispersal is 'optimal'.

The factors impacting the dispersal rate of RYMV throughout Madagascar were tested using a landscape phylogeographic approach (Dellicour, Rose, and Pybus 2016; Dellicour et al. 2016) similar to that recently applied to RYMV in mainland Africa (Dellicour et al. 2018b). Rice density, elevation, mean temperature, and annual precipitation are epidemiologically relevant and consequently of particular interest to test. For instance, absence of the disease in the very south of the country may link to the low rice density that restricts the possibilities of propagation. Sporadic presence of RYMV in the high-altitude plateau may be due to the lower temperature that is unfavorable to the virus vectors. Limited movements from west to east may be the consequence of elevation made by the high-altitude plateau that forms a physical barrier crossing the country from south to north. Yet, none of these factors were found to significantly impact the lineage dispersal velocity. This does not necessarily mean that they do not impact virus spread. For instance, if several of these factors interacted and concomitantly shaped the dispersal, their coordinated effect would not be identified in the seraphim analyses because only independent factors can be tested. Nonlinear relationships, such as between the spread and temperature because of detrimental effects of too low or too high values, may also be difficult to detect. Only linear relationships are investigated in the present study because of the lack of prior information on alternative relationships to test.

We also acknowledge that the temporal and/or the spatial scales of the factors we tested may not be epidemiologically relevant. Concerning the temporal scale, RYMV spread may be dependent on temperature variation over brief periods that cannot be captured by testing annual mean temperature. Concerning the spatial scale, a high degree of variation in rice cultivation over small areas in Madagascar that may be epidemiologically more relevant to assess host connectivity will be averaged out at a regional scale. Such small-scale variation is apparent in a suite of detailed geographical maps (http://www.cartomundi.fr/site/) that display rice cultivation in Madagascar in the 1970s at a scale of a few hundred meters (1/100,000). Areas of 10 × 10 km are often rice-free, even in regions with an overall high density of rice cultivation. These 10 × 10 km wide gaps are likely to be major obstacles for RYMV with vectors associated to restricted mobility and an average lineage dispersal rate of around 15 km/year. Moreover, the layout of the rice surfaces is critical. In particular, rice is often cultivated continuously along rivers and around lakes, facilitating rice connectivity for virus propagation, even across regions of overall low rice density. Altogether, this indicates that host density, even when considered at a scale of a few hundred meters, is not necessarily a good predictor of RYMV dispersal. These challenges may explain the difficulty to assess the drivers of the epidemics of RYMV in mainland Africa (Dellicour et al. 2018b) and now also in this study that can rely on detailed information for the more restricted island of Madagascar.

This study provides an example of data integration in viral phylodynamics (Baele et al. 2017), applied to a plant pathogen. The temporal signal in the sequence dataset of isolates from Madagascar was not sufficient to obtain a reliable estimate of the substitution rate although it included isolates collected over a 30-year-long period. However, a flexible partition model which shared the temporal information among several geographical partitions, yet allowing independent estimates of the other parameters, circumvented this difficulty. This enabled us to perform a phylogeographic reconstruction despite the low genetic diversity of the S4-Mg strain. Increasing the sequence length as an alternative way will not necessarily increase the temporal signal as recombination events now known to be frequent for RYMV may hamper the recovery of the clock signal. Increasing the sampling density may increase the spatial information and the temporal information. However, as illustrated by our study, it cannot account for extinct or displaced lineages which may be frequent in RYMV epidemiology because of the narrow host range of the virus. Using information from field surveys provided complementary spatial and temporal information on the early stages of the epidemic not accessible through genetic data. This study highlights the value and complementarity of (historical) nongenetic and (more contemporaneous) genetic surveillance data for reconstructing the history of spread of plant viruses. In conclusion, despite differences in strain characteristics, vector species, transmission means, landscape ecology, agro-climatic environments, and spatial scales, the virus dispersal processes across Madagascar and continental Africa are consistent.

Acknowledgements

We are grateful to Dr RAC Jones for helpful discussions. This work was partially funded by the French National Research Agency as an 'Investissements d'Avenir' program (ANR-10-LABX-001-01 Labex Agro) coordinated by Agropolis Foundation (1504-004 E-SPACE). B.V. was supported by the 'Fonds voor Wetenschappelijk Onderzoek' (FWO, Belgium). S.D. was supported by the 'Fonds Wetenschappelijk Onderzoek' (FWO, Belgium) and was currently funded by the 'Fonds National de la Recherche Scientifique' (FNRS, Belgium). The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (725422-ReservoirDOCS). P.L. acknowledges support by the Research Foundation—Flanders ('Fonds voor Wetenschappelijk Onderzoek—Vlaanderen' G066215N, G0D5117N, and G0B9317N). The funders had no role in the study design, data collection, and interpretation, or the decision to submit the work for publication. D.F. acknowledges the IRD i-Trop HPC at IRD Montpellier for providing HPC resources that have contributed to the research results reported within this article. URL: http://bioinfo.ird.fr/.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Conflict of interest: None declared.

References

Abo, M., Sy, A., and Alegbejo, M. (1997) '*Rice Yellow Mottle Virus* in Africa: Evolution, Distribution, Economic Significance and Sustainable Rice Production and Management Strategies', *Journal of Sustainable Agriculture*, 11: 85–111.

Abubakar, Z. et al. (2003) 'Phylogeography of *Rice Yellow Mottle Virus* in Africa', *The Journal of General Virology* , 84: 733–43.

Adego, A. et al. (2018) 'Full-Length Genome Sequences of Recombinant and Nonrecombinant Sympatric Strains of *Rice Yellow Mottle Virus* from Western Kenya', *Genome Announcements*, 6: 22.

Allarangaye, M. et al. (2006) 'Evidence of Non-Transmission of *Rice Yellow Mottle Virus* through Seeds of Wild Host Species', *Journal of Plant Pathology*, 88: 309–15.

Ayres, D. et al. (2012) 'BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics', *Systematic Biology*, 61: 170–3.

Baele, G. et al. (2017) 'Emerging Concepts of Data Integration in Pathogen Phylodynamics', *Systematic Biology*, 66: 47–65.

Bakker, W. (1974) 'Characterization and Ecological Aspects of *Rice Yellow Mottle Virus* in Kenya', *Agricultural Research Report*, 829: 1–152.

Baudelle, G., and Regnauld, H. (2004) *Echelles et temporalités en géographie*. Paris: SEDES, 176 pp.

Bourhy, H. et al. (2016) 'Revealing the Micro-Scale Signature of Endemic Zoonotic Disease Transmission in an African Urban Setting', *PLoS Pathogens*, 12: e1005525.

Chave, J. (2013) 'The Problem of Pattern and Scale in Ecology: What Have We Learned in 20 Years?', *Ecology Letters*, 16: 4–16.

Choi, J. Y. et al. (2019) 'The Complex Geography of Domestication of the African Rice *Oryza glaberrima*', *PLoS Genetics*, 15: e1007414.

Cliff, D., and Hagget, P. (1995) 'The Epidemiological Significance of Islands', *Health and Place*, 4: 199–209.

De Bruyn, A. et al. (2012) 'East African Cassava Mosaic-Like Viruses from Africa to Indian Ocean Islands: Molecular Diversity, Evolutionary History and Geographical Dissemination of a Bipartite Begomovirus', *BMC Evolutionary Biology*, 12: 228.

Dellicour, S. et al. (2016) 'SERAPHIM: Studying Environmental Rasters and Phylogenetically Informed Movements', *Bioinformatics*, 32: 3204–6.

—— et al. (2017) 'Using Viral Gene Sequences to Compare and Explain the Heterogeneous Spatial Dynamics of Virus Epidemics', *Molecular Biology and Evolution*, 34: 2563–71.

—— et al. (2018a) 'Phylodynamic Assessment of Intervention Strategies for the West African Ebola Virus Outbreak', *Nature Communications*, 9: 2222. doi 10.1038/s41467-018-03763-2.

—— et al. (2018b) 'On the Importance of Negative Controls in Viral Landscape Phylogeography', *Virus Evolution*, 4: 23.

——, Rose, R., and Pybus, O. (2016) 'Explaining the Geographic Spread of Emerging Epidemics: A Framework for Comparing Viral Phylogenies and Environmental Landscape Data', *BMC Bioinformatics*, 17: 12.

Delucchi, V. (2001) 'Biology and Control of *Dicladispa gestroi* Chapuis (Col., Chrysomelidae)', *Journal of Applied Entomology*, 125: 493–500.

Dijkstra, E. W. (1959) 'A Note on Two Problems in Connexion with Graphs', *Numerische Mathematik*, 1: 269–71.

Donque, G. (1975) *Contribution Géographique à l'Etude du Climat de Madagascar*. Tananarive: Nouvelle Imprimerie des Arts Graphiques à Tananarive, 477 pp.

Drummond, A. et al. (2003) 'Measurably Evolving Populations', *Trends in Ecology and Evolution*, 18: 481–8.

—— et al. (2006) 'Relaxed Phylogenetics and Dating with Confidence', *PLoS Biology*, 4: e88.

—— et al. (2012) 'Bayesian Phylogenetics with BEAUTY, and the BEAST 1.7', *Molecular Biology and Evolution*, 22: 1185–92.

Duchêne, S. et al. (2015) 'The Performance of the Date-Randomization Test in Phylogenetic Analyses of Time-Structured Virus Data', *Molecular Biology and Evolution*, 32: 1895–906.

Fargette, D. et al. (2008) '*Rice Yellow Mottle Virus*, an RNA Plant Virus, Evolves as Rapidly as Most RNA Animal Viruses', *Journal of Virology*, 78: 3584–9.

Faria, N. et al. (2012) 'Phylodynamics of the HIV-1 CRF02-AG Clade in Cameroon', *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 12: 453–60.

Fauquet, C., and Thouvenel, J.-C. (1977) 'Isolation of the *Rice Yellow Mottle Virus* in Ivory Coast', *Plant Disease Reporter*, 61: 443–6.

Firth, C. et al. (2010) 'Using Time-Structured Data to Estimate Evolutionary Rates of Double-Stranded DNA Viruses', *Molecular Biology and Evolution*, 27: 2038–51.

Gill, M. et al. (2013) 'Improving Bayesian Population Dynamics Inference: A Coalescent-Based Model for Multiple Loci', *Molecular Biology and Evolution*, 30: 713–24.

Gouy, M., Guindon, S., and Gascuel, O. (2010) 'SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building', *Molecular Biology and Evolution*, 27: 221–4.

Hébrard, E. et al. (2018) 'Identification of a Hypervirulent Pathotype of *Rice Yellow Mottle Virus*: A Threat to Genetic Resistance Deployment in West-Central Africa', *Phytopathology*, 108: 299–307.

Jacquot, M. et al. (2017) 'Bluetongue Virus Spread in Europe Is a Consequence of Climatic, Landscape and Vertebrate Host Factors as Revealed by Phylogeographic Inference', *Proceedings of the Royal Society B*, 284: doi 10.1098/rspb.2017.0919.

Jones, R. (2018) 'Plant and Insect Viruses in Managed and Natural Environments: Novel and Neglected Transmission Pathways', *Advances in Virus Research*, 101: 149–87.

Konaté, G., Sarra, S., and Traoré, O. (2001) '*Rice Yellow Mottle Virus* Is Seed-Borne but Not Seed Transmitted in Rice Seeds', *European Journal of Plant Pathology*, 107: 361–4.

King, A. et al. (2018) 'Changes to Taxonomy and the International Code of Virus Classification and Nomenclature Ratified by the International Committee on Taxonomy of Viruses', *Archives of Virology*, 163: 2601–31.

Le Bourdiec, F. (1977) 'Géographie Historique de la Riziculture Malgache', *Madagascar Revue de Géographie*, 31: 11–72.

Lemey, P. et al. (2010) 'Phylogeography Takes a Relaxed Random Walk in Continuous Space and Time', *Molecular Biology and Evolution*, 27: 1877–85.

Levin, S. (1992) 'The Problem of Pattern and Scale in Ecology: The Robert H. Mac Arthur Award Lecture', *Ecology*, 73: 1943–67.

Ling, R. et al. (2013) 'An Essential Fifth Codon ORF in the Sobemoviruses', *Virology*, 446: 397–408.

Lole, K. et al. (1999) 'Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination', *Journal of Virology*, 73: 152–60.

Ly, S. et al. (2018) 'Genetic Control of Seed Shattering during African Rice Domestication', *Nature Plants*, 4: 331–7.

Martin, D. et al. (2015) 'RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes', *Virus Evolution*, 1: 3.

McRae, B. H. (2006) 'Isolation by Resistance', *Evolution: International Journal of Organic Evolution*, 60: 1551–61.

—— et al. (2008) 'Using Circuit Theory to Model Connectivity in Ecology, Evolution and Conservation', *Ecology*, 89: 2712–24.

Monjane, A. et al. (2011) 'Reconstructing the History of Maize Streak Virus Strain a Dispersal to Reveal Diversification Hot Spots and Its Origin in Southern Africa', *Journal of Virology*, 85: 9623–36.

Murray, G. et al. (2016) 'The Effect of Genetic Structure on Molecular Dating and Tests for Temporal Signal', *Methods in Ecology and Evolution*, 7: 80–9.

Ndikumana, I. et al. (2017) 'Complete Genome Sequence of a New Strain of *Rice Yellow Mottle Virus* from Malawi, Characterized by a Recombinant VPg Protein', *Genome Announcements*, 5: 44.

N'Guessan, P. et al. (2001) 'Distribution, Pathogenicity, and Interactions of Two Strains of *Rice Yellow Mottle Virus* in Forested and Savanna Zones of West Africa', *Plant Disease*, 85: 59–64.

Ochola, D. et al. (2015) 'Emergence of *Rice Yellow Mottle Virus* in Eastern Uganda: Recent and Singular Interplay between Strains in East Africa and in Madagascar', *Virus Research*, 195: 64–72.

Peel, M., Finlayson, B., and McMahon, T. (2007) 'Updated World Map of the Köppen-Geiger Climate Classification', *Hydrology and Earth System Sciences*, 11: 1633–44.

Pinel, A. et al. (2000) 'Molecular Variability of Geographically Distinct Isolates of *Rice Yellow Mottle Virus*', *Archives of Virology*, 145: 1621–38.

Pinel-Galzi, A. et al. (2015) 'The Biogeography of Viral Emergence: *Rice Yellow Mottle Virus* as a Case Study', *Current Opinion in Virology*, 10: 7–13.

Portères, R. (1950) 'Vielles Agricultures de L'Afrique Intertropicale; Centres D'origine et de Diversification Variétale Primaire et Berceaux D'agriculture Antérieurs au XVIème Siècle', *Agronomie Tropicale*, 5: 489–507.

Pybus, O. et al. (2012) 'Unifying the Spatial Epidemiology and Molecular Evolution of Emerging Epidemics', *Proceedings of the National Academy of Sciences of the United States of America*, 109: 15066–71.

Rakotomalala, M. et al. (2008) 'Resistance to *Rice Yellow Mottle Virus* in Rice Germplasm in Madagascar', *European Journal of Plant Pathology*, 122: 277–86.

—— et al. (2013) '*Rice Yellow Mottle Virus* in Madagascar and in the Zanzibar Archipelago; Island Systems and Evolutionary Time Scale to Study Virus Emergence', *Virus Research*, 171: 71–9.

—— et al. (2014) 'First Report of *Rice Yellow Mottle Virus* in Rice in Federal Democratic Republic of Ethiopia', *Plant Disease*, 98: 1449.

Rambaut, A. et al. (2016) 'Exploring the Temporal Structure of Heterochronous Sequences Using TempEst (Formerly Path-O-Gen)', *Virus Evolution*, 2: vew007.

Reckhaus, P., and Andriamasintseheno, H. (1997) 'Rice Yellow Mottle Virus in Madagascar and Its Epidemiology in the Northwest of the Island', *Journal of Plant Diseases and Protection*, 104: 289–95.

——, and Randrianangaly, S. (1990) 'Rice Yellow Mottle Virus (RYMV) on Rice in Madagascar', *IRRI Newsletter*, 15: 30.

Rozas, J. et al. (2017) 'DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets', *Molecular Biology and Evolution*, 34: 3299–302.

Sarra, S., and Peters, D. (2003) 'Rice Yellow Mottle Virus Is Transmitted by Cows, Donkeys, and Grass Rats in Irrigated Rice Crops', *Plant Disease*, 87: 804–8.

Shapiro, B., Rambaut, A., and Drummond, A. (2006) 'Choosing Appropriate Substitution Models for the Phylogenetic Analysis of Protein Coding Sequences', *Molecular Biology and Evolution*, 23: 7–9.

Sõmera, M., Sarmiento, C., and Truve, E. (2015) 'Overview on Sobemoviruses and a Proposal for the Creation of the Family Sobemoviridae', *Viruses*, 7: 3076–115.

Suchard, M., and Rambaut, A. (2009) 'Many-Core Algorithms for Statistical Phylogenetics', *Bioinformatics*, 25: 1370–6.

Thompson, J., Higgins, D., and Gibson, T. (1994) 'CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice', *Nucleic Acids Research*, 22: 4673–80.

Tian, H. et al. (2018) 'Transmission Dynamics of Re-Emerging Rabies in Domestic Dogs of Rural China', *PLoS Pathogens*, 14: e1007392.

Traoré, O. et al. (2006) 'Rice Seedbed as Source of Primary Infection by *Rice Yellow Mottle Virus*', *European Journal of Plant Pathology*, 115: 181–6.

Trovão, N. et al. (2015a) 'Bayesian Inference Reveals Host-Specific Contributions to the Epidemic Expansion of Influenza a H5N1', *Molecular Biology and Evolution*, 32: 3264–75.

—— et al. (2015b) 'Host Ecology Determines the Dispersal Patterns of a Plant Virus', *Virus Evolution*, 1: vev016.

Uke, A. et al. (2015) 'Characterization of *Rice Yellow Mottle Virus* in North-Eastern Tanzania', *Journal of Agricultural Sciences*', *Tokyo University of Agriculture*, 60: 116–26.

Van Chi-Bonnardel, R. (1973) *Grand Atlas du Continent Africain*. Paris: Editions Jeune Afrique, 339pp.