Virus Research 195 (2015) 64–72

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Emergence of rice yellow mottle virus in eastern Uganda: Recent and singular interplay between strains in East Africa and in Madagascar

Dennis Ochola[a,1], Souley Issaka[b], Mbolarinosy Rakotomalala[c], Agnès Pinel-Galzi[d], Innocent Ndikumana[e], Judith Hubert[f], Eugénie Hébrard[d], Yacouba Séré[b,f], Geoffrey Tusiime[g], Denis Fargette[d,*]

[a] Cereals Research Program, National Crops Resources Institute (NaCRRI), P.O. Box 7084, Kampala, Uganda
[b] AfricaRice, 01 BP 2031, Cotonou, Benin
[c] CRRNO, FOFIFA, BP 289, Mahajanga, Madagascar
[d] Institut de Recherche pour le Développement (IRD), UMR RPB (IRD, CIRAD, Université Montpellier 2), Montpellier, France
[e] Crop Production Unit, Rwanda Agriculture Board, P.O. Box 5016, Kigali, Rwanda
[f] AfricaRice, P.O. Box 33581, Dar es Salaam, Tanzania
[g] Department of Crop Science, Faculty of Agriculture, Makerere University, Uganda

ARTICLE INFO

Article history:
Received 23 June 2014
Received in revised form
10 September 2014
Accepted 12 September 2014
Available online 20 September 2014

Keywords:
Virus emergence
Rice
Africa
Island biogeography

ABSTRACT

Epidemics of rice yellow mottle virus (RYMV) have developed recently in eastern Uganda, close to Lake Victoria in East Africa. Unexpectedly, all isolates from the affected area belonged to a single strain (named S4ug), a strain that is different from the S4lv strain that has been prevalent in the Lake Victoria basin for the past five decades. Interestingly, the S4ug strain is most closely related at the genomic level (except ORF1) to the strain present in Madagascar (S4mg), 2000 km away. The minor parent of the S4mg recombinant strain could not be detected. Molecular clock dating analysis indicated that the singular sequence of events – that associated the emergence of a new strain (S4ug), a modular recombination between closely related strains (S4mg and S4ug) and a long distance transmission (S4mg) – occurred recently, within the past few decades. This finding is at variance with the process of gradual strain dispersal and diversification over two centuries throughout Africa that was previously established.

1. Introduction

Rice yellow mottle virus (RYMV) was first detected in 1966 at Kisumu in Western Kenya near Lake Victoria (Bakker, 1974), close to the border with eastern Uganda. RYMV has since been observed and detected in almost all rice-producing countries of sub-Saharan Africa (Abo et al., 1998). 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 other wild *Poaceae* species (Bakker, 1974). Field studies have revealed the significant role of rice nurseries in the propagation of RYMV (Sarra et al., 2004; Traoré et al., 2006). Movement of the virus over short distances by mammals (Sarra and Peters, 2003) and by birds (Peters et al., 2012) has been reported. Like several other sobemoviruses, RYMV is transmitted by beetle species of the family Chrysomelidae (Bakker, 1974), which generally have a restricted dispersal range. Transmission by grasshoppers, insects with stronger flight ability, has been postulated (Bakker, 1974) but not established experimentally for RYMV or for any other sobemovirus. RYMV is not transmitted through rice seeds or through seeds of wild host species (Abo et al., 2004; Allarangaye et al., 2006; Fauquet and Thouvenel, 1977; Konaté et al., 2001). Overall, there is no documented means of long-distance movement > 100 km. The clear-cut spatial pattern of diversity and the significant relationships between the geographic and the genetic distances (Abubakar et al., 2003), characteristic of an isolation-by-distance model, implies that RYMV spread is generally gradual. Study of the colonisation of islands at various distances from mainland Africa revealed that RYMV dispersal over distances of several dozens of kilometres is recurrent whereas that over hundreds of kilometres is possible but exceptionally rare (Rakotomalala et al., 2013).

Rice yellow mottle virus is a single-stranded RNA species of the genus *Sobemovirus* with five open reading frames (ORF) (Truve and Fargette, 2011; Ling et al., 2013). ORF1, located at the 5′ end

* Corresponding author. Tel.: +33 467416460.
E-mail address: denis.fargette@ird.fr (D. Fargette).
[1] Present address: Commodity Systems and Genetic Resources Program, Bioversity International, P.O. Box 24384, Kampala, Uganda.

http://dx.doi.org/10.1016/j.virusres.2014.09.004
0168-1702/

of the genome, encodes a small protein involved in virus movement (Bonneau et al., 1998) and in gene silencing suppression (Siré et al., 2008; Lacombe et al., 2010). ORF2, which encodes the central polyprotein, has two overlapping ORFs. ORF2a encodes a serine protease and a viral genome-linked protein involved in virulence (Hébrard et al., 2010). 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).

RYMV is highly variable with a well defined geographical distribution of the genetic variants in mainland Africa. The greatest diversity has been found in East Africa (Pinel et al., 2000). Successive surveys in Tanzania (Kanyeka et al., 2007), and the first surveys in the neighbouring countries of Rwanda (Ndikumana et al., 2011), Burundi (Ndikumana et al., 2012) and Democratic Republic of Congo (Hubert et al., 2013) showed strain circulation within this region but no emergent strains. In Madagascar, all isolates belong to a single strain (S4mg); it is most likely that this introduction was a singular event (Rakotomalala et al., 2013). The distant genetic relationships of the S4mg strain with the other strains does not provide clear evidence on the geographical origin of the introduction. Analysis of the diversity in eastern Tanzania revealed recombination events between isolates of closely related strains (Pinel-Galzi et al., 2009; Rakotomalala et al., 2013).

Recent epidemics in eastern Uganda occurred alongside Lake Victoria where the disease incidence exceeded 70% (Ochola and Tusiime, 2011). Samples representative of the area affected by the new epidemic were collected. The capsid protein gene of RYMV in these samples was amplified and sequenced. They were analysed with isolates of the prevailing strain in the Lake Victoria basin, and with a set of isolates identified in 21 countries in Africa between 1966 and 2012. In addition, isolates from the new epidemics were fully sequenced and compared to a set of isolates representative of the geographical distribution and of the genetic diversity of the virus in Africa and in Madagascar. As RYMV is a measurably evolving population (Fargette et al., 2008), the age of the different strains was estimated. It was concluded that the recent epidemics in eastern Uganda were caused by an emergent strain. Its unexpected characteristics and its singular relationships with other strains in East Africa and in Madagascar challenged previous views on the evolution and the epidemiology of RYMV.

2. Materials and methods

2.1. Samples

Three field surveys were conducted in 2009 and 2010 in 11 districts of eastern Uganda between Lake Kyoga (1°30′ North, 33° South) and the Kenyan border at the north of Lake Victoria. The surveys covered an area of 200 km east–west × 100 km north–south (Ochola and Tusiime, 2011). Samples representative of the area affected were collected. The ORF4 of RYMV (720 nt) in 26 samples was amplified and sequenced as described previously (Pinel et al., 2000) and compared to 419 RYMV sequences of isolates collected from 1966 to 2012 in 21 African countries (Table S1). It included 32 isolates of the prevailing local strain – referred to as the S4lv strain – sampled alongside the Lake Victoria basin in southern Uganda (two isolates), western Kenya (3), Burundi (2), eastern Democratic Republic of Congo (3), northern Tanzania (15), and Rwanda (7). Four isolates from Uganda were fully sequenced (ca. 4450 nt) as described previously (Fargette et al., 2004), including three isolates from the recent epidemics and one isolate from Uganda that was collected in 2000 in the south of the country (Pinel-Galzi et al., 2006). They were compared to a set of 33 isolates representative of the geographical distribution and of the genetic diversity of RYMV. Altogether, 37 full-length sequences were analysed (Supplementary Table S1).

2.2. Genetic and phylogenetic analyses

The ORF4 sequences of the isolates were aligned using CLUSTAL X with default parameters (Thompson et al., 1994). The diversity index (PI) which is the average number of nucleotide substitutions per site between two sequences of each individual population, and the average number of nucleotide substitutions per site between populations (Dxy) were calculated using DNAsp (http://www.ub.edu/dnasp/; Librado and Rozas, 2009). The General Time Reversible Model with a rate variation and a class for invariant sites I (GTR + G4 + I) was selected as the best-fit nucleotide substitution model by a Modeltest selection procedure (Posada and Crandall, 1998) accessed through the Datamonkey interface (http://www.datamonkey.org). The maximum-likelihood (ML) phylogenetic tree was inferred using the PHYML 3.0 algorithm implemented in the SEAVIEW 4.5 software (Gouy et al., 2010; Guindon et al., 2010; http://doua.prabi.fr/software/seaview). The ML phylogenetic tree of the ORF4 sequences was rooted at the point in the tree that minimises the variance of root-to-tip distances. To assess the reliability of key nodes of the trees, a bootstrap resampling procedure was used with 100 replicate trees estimated under the substitution model. The full-length sequences of the isolates were aligned and the phylogenetic tree was reconstructed similarly.

2.3. Molecular clock dating

The Time of the Most Recent Common Ancestor (TMRCA) of the nodes of interest and their associated 95% highest density probability (HPDs) intervals were estimated within a Bayesian coalescent framework by a Markov Chain Monte-Carlo (MCMC) implemented in BEAST v1.7.5 (http://beast.bio.ed.ac.uk) (Drummond et al., 2006; Drummond and Rambaut, 2007). A relaxed (uncorrelated lognormal) clock model, a GTR + G4 + I evolutionary model, and the flexible Bayesian skyline population model as a coalescent prior were selected. In all cases, the BEAST analyses were run until all relevant parameters converged (ESS > 100) with 20% of the MCMC chains discarded as burn-in. Statistical confidence in the parameter estimates was represented by values for the 95% HPDs intervals around the marginal posterior parameter means.

2.4. Detection of recombinants

The search for recombinant signals was conducted in the 37 full-length sequence alignment. Recombination events were compared to those detected in earlier analyses of 27 and 33 full-length sequences (Pinel-Galzi et al., 2009; Rakotomalala et al., 2013), i.e., without the sequences of the isolates from eastern Uganda. The recombination events were assessed and delineated with the RDP4 version 4.34 software (http://darwin.uvigo.es/rdp/rdp.html) (Martin et al., 2010). RDP4 incorporates seven recombination-detecting algorithms: BOOTSCAN, CHIMAERA, GENECONV, MAXCHI, RDP, SISCAN, and 3SEQ. These analyses were performed using default settings from the different detection programmes and a Bonferroni corrected *P*-value cut-off of 0.01. The recombination signals that were detected by five or more of the seven methods implemented in RDP4 with the default parameters were considered. The second criterion to assess the strength of the recombination signal was the overall corrected *P*-value of the