Virus Research 238 (2017) 193–197

Contents lists available at ScienceDirect

Virus Research

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

Short communication

Novel mastreviruses identified in Australian wild rice

Simona Kraberger[a], Andrew D.W. Geering[b,**], Matthew Walters[c], Darren P. Martin[d], Arvind Varsani[a,c,e,*]

[a] The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ 85287-5001, USA
[b] Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, 41 Boggo Road, Dutton Park, QLD 4102, Australia
[c] School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand
[d] Computational Biology Group, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa
[e] Structural Biology Research Unit, Department of Clinical Laboratory Sciences, University of Cape Town, Observatory, Cape Town, South Africa

ARTICLE INFO

Keywords:
Geminivirus
Oryza sp. 'Taxon A', 'Taxon B'
Oryza australiensis

ABSTRACT

Most known mastreviruses (family *Geminiviridae*) infect members of the grass family, Poaceae. Although the greatest number of grass-infecting mastrevirus species have been discovered in Africa, it is apparent that the ten grass-infecting mastrevirus species that have so far only been discovered in south-east Queensland have a degree of diversity that rivals that observed in Africa. In this study, we have used a deep sequencing approach to identify two new mastrevirus species, tentatively named rice latent virus 1 and 2 (RLV 1 and 2), from two, undescribed wild rice species (*Oryza* AA genome group) in Cape York Peninsula, Queensland. The sequences of these new viruses had less than 70% identity with any previously identified mastrevirus, and therefore their discovery vastly expands the known diversity of monocot-infecting mastreviruses in Australia. This study also highlights the potential risks of novel crop pathogens emerging from uncultivated grass species, as the wild rice hosts are very closely related to domesticated rice.

Various mastreviruses (genus *Mastrevirus*, family *Geminiviridae*) are pathogens of economically important crops, most notably maize in Africa and chickpea in the Middle East. Thirty-seven mastrevirus species have been documented from Africa, the Americas, Asia, Australia, Europe, and various Indian and Pacific Ocean islands. Seven of these known species infect dicotyledonous plants and 30 species infect monocotyledonous plants (Agindotan et al., 2015; Candresse et al., 2014; Kraberger et al., 2015; Muhire et al., 2013; Oluwafemi et al., 2014; Zerbini et al., 2017). The monocot-infecting mastreviruses are the most widely distributed, genetically diverse group, and infect both cultivated and wild members of the Poaceae (Kraberger et al., 2013; Muhire et al., 2013; Oluwafemi et al., 2014; Varsani et al., 2009). Africa and Australia have been identified as global diversity hotspots for mastreviruses (Kraberger et al., 2013, 2012; Varsani et al., 2008), containing about 65% of the known species.

Rice (*Oryza sativa*) is a staple crop for a large proportion of the global population with more than 480 million metric tons being produced annually (Muthayya et al., 2014). The genus *Oryza* comprises 22–24 species, which are pantropical in distribution and divided into six diploid (A–C and E–G) and five tetraploid (BC, CD, HJ, HK and KL) genome groups, based on molecular systematics and reproductive compatibility (Ammiraju et al., 2010; Ge et al., 1999; Lu et al., 2009). *O. sativa*, and the only other cultivated rice species, *O. glaberrima*, are both classified as having an AA genome. Australia is a centre of diversity and the possible origin of the AA genome group (Brozynska et al., 2017). There is considerable scientific interest in the wild species because of their potential to be used in rice breeding programs (Henry et al., 2010) but also because of the risks they pose to rice production by acting as reservoirs of pests and pathogens (Khemmuk et al., 2016; Petrovic et al., 2013). The fungus *Pyricularia oryzae*, the cause of blast disease, is thought to have jumped from wild to domesticated rice in north Queensland, an event that has led to the temporary abandonment of cropping in parts of Cape York Peninsula until disease management strategies can be devised (Khemmuk et al., 2016).

Surveys for monocot-infecting mastreviruses have predominantly been in south-east Queensland, close to the major urban centre, Brisbane. The vast monsoonal grasslands of northern Australia have never been explored for mastreviruses, mainly due to logistical reasons. The wild rice species found in Cape York Peninsula, Queensland (*Oryza australiensis* (EE genome) and two undescribed AA genome-type species,

* Corresponding author at: The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life sciences, Arizona State University, Tempe, AZ, 85287-5001, USA.
** Corresponding author at: Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, 41 Boggo Road, Dutton Park, QLD 4102, Australia.
E-mail addresses: Andrew.Geering@daf.qld.gov.au (A.D.W. Geering), Arvind.varsani@asu.edu (A. Varsani).

http://dx.doi.org/10.1016/j.virusres.2017.07.003
Received 19 May 2017; Received in revised form 29 June 2017; Accepted 2 July 2017
Available online 04 July 2017
0168-1702/

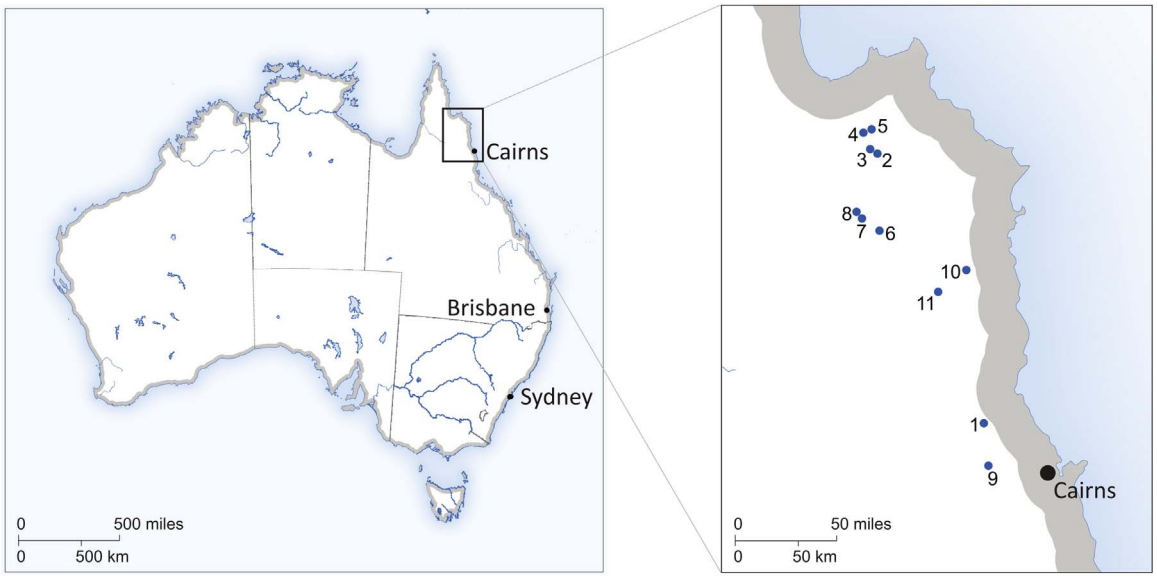

Fig. 1. Location of wild rice sampling sites in the Cape York Peninsula, Queensland, Australia. Fifty-three samples were collected from 11 different sites (see Table 1 for sample and site details). Rice latent virus 1 was identified in samples from ten sites (1–10) where either *Oryza* sp. 'Taxon A' or 'Taxon B' was growing. Rice latent virus 1 was only identified at site 5. At site 11, *Oryza australiensis* was present and this was the only site where no mastreviruses were detected.

Oryza sp. Taxon A and *Oryza* sp. Taxon B), grow on flood plains and the fringes of lagoons, and are perennial, dying back to buried rhizomes during the dry season (Henry et al., 2010). Many perennial monocots, such as sugarcane and banana, support large assemblages of plant viruses, probably due to the fact that infections persist from one generation to the next via vegetative propagation of the plants e.g. potyviruses (Xu et al., 2008). In this study, we hypothesized that the wild rice species in Queensland would likewise be hosts of plant viruses, especially mastreviruses, given the long presence of this group of viruses in Australia.

During 12–14 May 2014 and 27–29 March 2015, wild rice plants at various sites (n = 11) in Cape York Peninsula, from the Atherton Tableland to Rinyirru (Lakefield) National Park (Fig. 1; Table 1), were sampled for testing for leaf spot pathogens (see Khemmuk et al. (2016)) and viruses (this study). Genomic DNA extracts from 58 wild rice plants (Table 1) were prepared using a GF-1 Nucleic Acid Extraction Kit (Vivantis Technologies, Malaysia). One µl of genomic DNA from each sample was used to amplify circular DNA by rolling circle amplification using an Illustra TempliPhi Amplification Kit (GE Healthcare, USA). The RCA products were pooled (1 µl from each amplification) and sequenced on an Illumina HiSeq 2500 platform at Macrogen Inc (Korea). Resulting paired-end reads were *de novo* assembled using ABySS 1.9 (Simpson et al., 2009) and contigs > 1000 nt were analysed against a local viral database using BLASTX (Altschul et al., 1990). Two contigs (~2.7 kb) were identified that potentially encode proteins with detectable similarities to those of mastreviruses. Pairs of abutting primers were designed (RLV-1F: 5′- CCC ACC GCA CAA TAC AAT GTT ATA GAC-3′ and RLV-1R: 5′- GTC ATA AGT GGT GAA GTC CAC CAG-3′; RLV-2F: 3′-TGT GTT TCT GAG CTT CAT AAA GAT GGG-3′ and RLV-2R: 5′- TAA AAC GCC GTA AGG CCT GAG CAG A-3′) to screen and recover complete genomes of these mastreviruses from each of the wild rice samples. PCR was done using HiFi hotstart DNA polymerase (Kapa Biosystems, USA) and 1 µl of RCA products as a template in a 20 µl total volume. Thermal cycling conditions were: 94 °C for 3 min, 25 cycles of 98 °C (20 s), 55 °C (30 s), 72 °C (3 min) with a final extension of 72 °C for 3 min. PCR products were resolved on a 0.7% agarose gel and fragments ~2.7 kb in length excised and gel purified. The DNA was then cloned in pJET1.2 plasmid (Thermo Fisher Scientific, USA) and Sanger sequenced (Macrogen Inc., Korea) by primer walking. From 22 of the 58 wild rice samples, mastrevirus genomes were recovered using primer pair RLV-1F/R and from one additional sample another genome was recovered using primer pair RLV-2F/R (Table 1).

Viral sequences were assembled using DNABaser (Heracle BioSoft S.R.L., Romania) and open reading frames identified using DNAMAN V7 (Lynnon Biosoft, Canada). Datasets including species and strains from all previously identified mastreviruses (downloaded on 25th April 2017) were compiled in MEGA V7 (Kumar et al., 2016), aligned using MUSCLE (Edgar, 2004) and a full genome neighbor-joining phylogenetic tree was constructed using Jukes-Cantor substitution model, 1000 bootstrap replicates, and rooted with beet curly top Iran virus sequences (Fig. 2). Maximum likelihood phylogenetic trees were constructed for CP and Rep amino acid sequence datasets using PhyML3 (Guindon et al., 2010) implemented in Seaview V4 (Gouy et al., 2010) with aLRT support, and Eragrostis curvula streak virus to root the trees. Best fit models for CP (LG + I + G + F) and Rep (LG + I + G) were identified using ProtTest (Darriba et al., 2011) (Fig. 2).

Pairwise comparison of the full genome dataset using SDT v1.2 (Muhire et al., 2014) indicated that all new sequences had less than 70% pairwise identity with all other known mastreviruses (Supplementary data 1). Following the guidelines for mastrevirus species demarcation (i.e. a species demarcation threshold of 78% coupled with phylogenetic support (Muhire et al., 2013), two novel mastrevirus species were identified, one represented by 22 virus isolates and the other, by a single isolate. No wild rice specimen had overt disease symptoms such as foliar chlorotic mosaics or streaking and therefore the two new mastrevirus species were tentatively named rice latent virus 1 (n = 22) and 2 (n = 1) (RLV 1 and 2). RLV 1 was detected in both *Oryza* sp. 'Taxon A' and 'B' but not *O. australiensis*, although only a single location containing the latter was surveyed. RLV 1 was very common and widely distributed, as it was found at nearly every site that was visited. However, RLV 2 was only found at one site, Red Lily Lagoon in Rinyirru National Plant, in a single plant of *Oryza* sp. 'Taxon B'.

The 22 RLV 1 genomes share > 98% identity with one another and range in size from 2754 to 2758 nt. The single RLV 2 genome is 2843 nt in length. Both viruses had typical mastrevirus genome architecture with genes encoding a movement protein (MP) and a capsid protein (CP) on the virion-sense strand and a RepA and a replication associated protein (Rep) on the complementary sense-strand. The nonanucleotide motif at the presumed origin of virion-strand replication in RLV 2 is identical to that of most other monocot-infecting mastreviruses (TAATATTAC), whereas RLV 1 has a TAATGTTAC sequence that has only previously been observed in the genomes of some dicot-infecting