

# Genome sequence analysis of two South African isolates of *Moroccan watermelon mosaic virus* infecting cucurbits

Jacques Davy Ibaba<sup>1</sup> · Mark D. Laing<sup>1</sup> · Augustine Gubba<sup>1</sup>

Received: 19 February 2016 / Accepted: 7 July 2016 / Published online: 23 July 2016  
© Springer Science+Business Media New York 2016

**Abstract** *Moroccan watermelon mosaic virus* (MWMV) has been prevalent in cucurbits in the Republic of South Africa (RSA) since it was first reported in 1987. However, full genome studies of the South African isolates have never been conducted previously. The full genome of two MWMV isolates infecting cucurbits (*Cucurbita pepo* L.) in the province of KwaZulu-Natal, RSA, was compared with the genome of the Tunisian isolate in this communication. The genome sequences of the RSA MWMV isolates were elucidated using next-generation sequencing and Sanger sequencing. The analyses performed included nucleotide and amino acid sequence comparison, determination of the genetic distances, detection of potential recombination, and phylogeny. The genome sequences of the RSA MWMV isolates were found to be 9719 nucleotides long, excluding the poly(A) tail. Sequence homology, genetic distances, and phylogenetic analyses indicated close relationships between the RSA isolates. This record will contribute to building up the MWMV isolate sequences from the different countries where the virus occurs, a useful step toward understanding MWMV evolution.

**Keywords** *Potyvirus* · Phylogenomics · ssRNA virus · Next-generation sequencing · Southern Africa · Cucurbit

Edited by Seung-Kook Choi.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11262-016-1372-4) contains supplementary material, which is available to authorized users.

✉ Augustine Gubba  
gubbaa@ukzn.ac.za

<sup>1</sup> Department of Plant Pathology, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Scottsville, Pietermaritzburg, South Africa

Virus diseases have impeded cucurbit cultivation in the Republic of South Africa (RSA) [1, 2]. Viruses, especially from the genus *Potyvirus* in the family *Potyviridae*, can cause severe damage to cucurbits, with losses up to 100 % when infection occurs early in the season [3]. *Moroccan watermelon mosaic virus* (MWMV), a potyvirus infecting cucurbits, was reported for the first time in Morocco in 1974 as a strain of *Watermelon mosaic virus* (WMV) on the basis of its host range [4]. It was later found to be a distinct potyvirus based on coat protein (CP) sequence analysis with other potyviruses [5]. At a molecular level, MWMV is part of the *Papaya ringspot virus* (PRSV) cluster [4]. MWMV was reported in cucurbits cultivated in RSA in 1987 [2]. Symptoms associated with MWMV on cucurbits include mosaic, stunting, dark-green blisters, vein banding, and leaf and fruit malformation [2, 4]. MWMV has been identified, in RSA, as the most prevalent potyvirus infecting cucurbits in various surveys [1, 2, 6]. MWMV has also been reported to occur in Zimbabwe [4], Tanzania [7], DRC [8], Cameroon [4], Nigeria [9], Niger [4], Sudan [10], Canary Island [4], France [11], Greece [12], Italy [13], Portugal [4], and Spain [4]. Most molecular studies on the MWMV genome have been centered on the portion of the genome flanking the coding sequences of the C-terminal part of the Nib to the N-terminal part of the CP [1, 8–13]. Only one complete genome sequence of MWMV, a Tunisian isolate (accession number: EF579955) [4], is available on the NCBI GenBank. The full genome sequence of two MWMV isolates infecting cucurbits cultivated in the province of KwaZulu-Natal (KZN), RSA were elucidated and compared with the Tunisian isolate in this communication.

Baby marrow (*Cucurbita pepo* L.) and pattypan (*C. pepo* L.) were identified as highly susceptible to MWMV during the surveys that were conducted in the cucurbit growing areas of KZN in the 2011, 2012, and 2013

growing seasons [1]. Consequently, two MWMV-infected leaf samples, from a pattypan and a baby marrow, were randomly selected for next-generation sequencing (NGS) to recover the full genome of MWMV. Total RNA, extracted using the plant RNA NucleoSpin kit (Macherey–Nagel, Germany), was treated with Ribo-Zero Plant prior to NGS library preparation and 125 × 125 bp paired-end sequencing on an Illumina HiSeq. The library preparation and NGS were performed at the Agricultural Research Council’s Biotechnology Platform (ARC-BTP) in Pretoria (RSA). FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to assess the quality of the reads before and after trimming and removal of adapters that were performed using Trimmomatic (version 0.33) software [14]. De novo assembly was subsequently performed using SeqMan NGen (version 12.3.1 build 48; DNASTAR) software with the default settings. De novo assembly was chosen over the option of mapping reads to a reference genome as a precautionary measure in case the genome of the RSA MWMV isolates had undergone recombination, or considerably diverged from the genome sequence of the Tunisian isolate. The host genome data was excluded during *de novo* assembly. The contigs generated were blasted to the GenBank database using the Netsearch function on the SeqMan pro (version 12.3.1 build 48 421; DNASTAR) software.

The longest MWMV identified contig in the pattypan sample was 9730 bp. MWMV mapped reads from the pattypan sample were 3,453,027 (Median Coverage: 42054.34), and had an average length of 114 bp. Regarding the MWMV mapped reads from the baby marrow sample, the longest contig identified was 9735 bp, made up of 3,425,094 reads (Median Coverage: 37626.61) with an

average length of 113 bp. A gap from position 6 to 11 on the genome of the Tunisian isolate was the only ambiguity spotted when aligning the genome sequences of the RSA with the Tunisian MWMV isolates. This gap was found to be consistent after direct sequencing of the RT-PCR amplicon obtained using the primers MWMVF1: 5′-AAA CAC TCA ACA CAA CAC AAC ATC-3′, and MWMVR530: 5′-CCC TGT CTT GCT TCA GCT AAA TTC-3′. Although 5′ RACE experiments to determine the accurate 5′ ends sequence of the isolates under study were not carried out, further Sanger sequencing of the 5′ end sequences obtained using a Double-stranded cDNA synthesis kit (ThermoScientific, USA) yielded the same sequences as those generated *de novo*. This led us to conclude that *de novo* assembly gave an accurate sequence of the 5′ ends of the RSA isolates. The 9.73 kb contigs from the baby marrow and pattypan samples were therefore considered as the consensus sequences of the genome of the RSA MWMV isolates. Both genome sequences of the RSA MWMV isolates (accession number: KU315175 and KU315176) consist of 9719 nucleotides, excluding the poly(A) tail, organized into 150 noncoding nucleotides at its 5′ terminus followed by 9375 nucleotides encoding the polyprotein, and 194 noncoding nucleotides at 3′ terminus (Table 1). The two RSA isolates share the same predicted polyprotein cleaving sites as the Tunisian isolate.

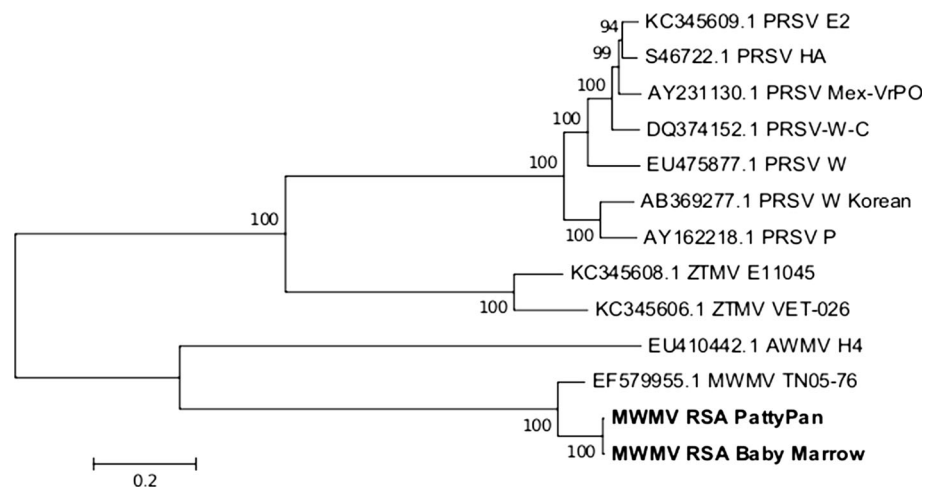
Comparative genomics of the RSA and Tunisian MWMV isolates included nucleotide and amino acid sequence comparisons, evaluation of genetic distances, and phylogenetic analysis. The online tool SIAS (<http://imed.med.ucm.es/Tools/sias.html>) was used for sequence comparisons. Genetic distances and phylogenetic analyses were performed using MEGA6 software [15]. Multiple sequence

**Table 1** Genome organization of SA isolates of MWMV and protein MW comparison

Genome position (nt)	Genome features	Protein MW (kDa)		
		Baby marrow isolate	Pattypan isolate	Tunisian isolate
1–150	5′ noncoding region	–	–	–
151–9525	Polyprotein ORF	354.06	354.07	354.57
9526–9719	3′ noncoding region	–	–	–
15–1185	P1	38.69	38.68	39.19
1186–2556	HC-Pro	51.63	51.65	51.63
2557–3597	P3	39.71	39.75	39.58
3598–3753	6K1	6.05	6.05	6.05
3754–5658	CI	71.06	71.1	71.1
5659–5829	6K2	6.61	6.61	6.71
5830–6399	VPg	21.53	21.57	21.49
6400–7116	NIa	26.92	26.92	26.86
7117–8667	NIb	59.50	59.47	59.43
8668–9525	CP	32.52	32.42	32.44
3018–3200	PIPO	6.82	6.82	6.87

–: Not applicable, *nt* nucleotide

**Fig. 1** Maximum likelihood Phylogram of the Moroccan watermelon mosaic virus (MWMV) isolates in the Papaya ringspot virus (PRSV) cluster. The bootstrap values are indicated next to the branches. The branch lengths are measured in the number of substitutions per site. ZTMV: *Zucchini tigré mosaic virus* [18], AWMV: *Algerian watermelon mosaic virus* [19]



alignment was performed using the Muscle program, implemented in MEGA6 software. The genetic distances were computed using the maximum composite likelihood model [16] with a gamma distribution. The phylogeny of the MWMV isolates was inferred with 500 iterations bootstrap analysis using the maximum likelihood method based on the general time reversible model with a discrete gamma distribution and evolutionarily invariable sites. Putative recombination events on the genomes of the RSA isolates were checked using the automated methods in the RDP (version 4.56) software [17].

The genome sequences of the RSA MWMV isolates were almost identical to each other (Table S1). A 100 % nucleotide sequence identity was observed between the RSA isolates for the 5' and 3' noncoding regions (NCR), 6K1 and PIPO. The lowest nucleotide sequence identity, *i.e.*, 98.93 %, was recorded for P1. The nucleotide sequence identity of the RSA MWMV isolates varied between 80 and 95 % (Table S1) when compared with the Tunisian isolate, with the highest nucleotide sequence identity recorded with PIPO, followed by the 3' NCR. The 5' NCR had the lowest nucleotide sequence identity among all MWMV isolates. In terms of the amino acid sequence identity (Table S1), a 100 % sequence identity was observed between the RSA MWMV isolates for the 6K1, 6K2, NIa, and PIPO. The amino acid sequence identities for the VPg and P1 were the lowest at around 98 %. The RSA and the Tunisian MWMV isolates share between 82 and 99 % amino acid sequence identity.

RSA MWMV isolates were expected to be less diverse from each other than the Tunisian isolate. The genetic distances (Table S2) confirmed this hypothesis. Genetic distances between the RSA isolates were zero for PIPO and the 6K1. The longest distance was observed for P1. Between the RSA and the Tunisian isolates, the genetic distances were as follows in ascending order: PIPO, P3, CI, NIa, N1b, 6K1, VPg, HC-Pro, 6K2, P1, and CP. The

highlight of this test was the high level of divergence between the CP of the RSA and Tunisian MWMV isolates. This diversity is consistent with the previous studies conducted by Yakoubi *et al.* [4] and Ibaba *et al.* [1]. No recombination event was identified in the genome of all MWMV isolates. Phylogenetic tree (Fig. 1) showed that all MWMV isolates formed a monophyletic group sharing a common ancestor with the *Algerian watermelon mosaic virus* isolate H4 (EU410442.1). This study provided two genome sequences of the RSA MWMV isolates infecting cucurbits. Future genome studies of MWMV isolates from the other countries where it occurs may shed light on the evolution of this virus.

**Acknowledgments** The authors are grateful to Mr Jonathan Featherston and Ms Thulile Faith Nhlapo, from the ARC-BTP, for their technical guidance on NGS that they provided. Sanger Sequencing was performed at Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, RSA).

#### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

#### References

- J.D. Ibaba, M.D. Laing, A. Gubba, *Crop. Prot.* **75**, 46–54 (2015). doi:[10.1016/j.cropro.2015.04.019](https://doi.org/10.1016/j.cropro.2015.04.019)
- F.W. van der Meer, H.M. Garnett, *J. Phytopathol.* **120**, 255–270 (1987). doi:[10.1111/j.1439-0434.1987.tb04440.x](https://doi.org/10.1111/j.1439-0434.1987.tb04440.x)
- H. Lecoq and C. Desbiez, in *Adv. Virus Res.*, edited by L. Gad L. Hervé (Academic Press, 2012), pp. 67–126. doi:<http://dx.doi.org/10.1016/B978-0-12-394314-9.00003-8>
- S. Yakoubi, C. Desbiez, H. Fakhfakh, C. Wipf-Scheibel, M. Marrakchi, H. Lecoq, *Arch. Virol.* **153**, 117–125 (2008). doi:[10.1007/s00705-007-1074-2](https://doi.org/10.1007/s00705-007-1074-2)

5. N.M. McKern, P.M. Strike, O.W. Barnett, C.W. Ward, D.D. Shukla, *Arch. Virol.* **131**, 467–473 (1993). doi:[10.1007/BF01378647](https://doi.org/10.1007/BF01378647)
6. K.R. Craddock, M.D. Laing, J.V. da Graça, *Rev. Mex. Fitopatol.* **19**, 251–252 (2001). ([articulo.oa?id = 61219220](https://doi.org/10.1007/s00705-011-1124-7))
7. W. Menzel, M. Abang, S. Winter, *Arch. Virol.* **156**, 2309–2311 (2011). doi:[10.1007/s00705-011-1124-7](https://doi.org/10.1007/s00705-011-1124-7)
8. Y. Arocha, N. Vigheri, B. Nkoy-Florent, K. Bakwanamaha, B. Bolomphety, M. Kasongo, P. Betts, W.A. Monger, V. Harju, R.A. Mumford, P. Jones, *Plant. Pathol.* **57**, 387 (2008). doi:[10.1111/j.1365-3059.2007.01658.x](https://doi.org/10.1111/j.1365-3059.2007.01658.x)
9. A.T. Owolabi, F. Rabenstein, F. Ehrig, M. Maiss Edgar, H.J. Vetten, *Int. J. Virol.* **8**, 258–270 (2012). doi:[10.3923/ijv.2012.258.270](https://doi.org/10.3923/ijv.2012.258.270)
10. H. Lecoq, G. Dafalla, C. Desbiez, C. Wipf-Scheibel, B. Delecolle, T. Lanina, Z. Ullah, R. Grumet, *Plant. Dis.* **85**, 547–552 (2001). doi:[10.1094/pdis.2001.85.5.547](https://doi.org/10.1094/pdis.2001.85.5.547)
11. H. Lecoq, I. Justafre, C. Wipf-Scheibel, C. Desbiez, *Plant. Pathol.* **57**, 766 (2008). doi:[10.1111/j.1365-3059.2008.01848.x](https://doi.org/10.1111/j.1365-3059.2008.01848.x)
12. I. Malandraki, N. Vassilakos, C. Xanthis, G. Kontosfiris, N.I. Katis, C. Varveri, *Plant. Dis.* **98**, 702 (2013). doi:[10.1094/PDIS-10-13-1100-PDN](https://doi.org/10.1094/PDIS-10-13-1100-PDN)
13. P. Roggero, G. Dellavalle, V. Lisa, V.M. Stravato, *Plant. Dis.* **82**, 351 (1998). doi:[10.1094/PDIS.1998.82.3.351B](https://doi.org/10.1094/PDIS.1998.82.3.351B)
14. A.M. Bolger, M. Lohse, B. Usadel, *Bioinformatics* **30**, 2114 (2014). doi:[10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170)
15. K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, *Mol. Biol. Evol.* **30**, 2725–2729 (2013). doi:[10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197)
16. K. Tamura, M. Nei, S. Kumar, *Proc. Natl. Acad. Sci. USA.* **101**, 11030–11035 (2004). doi:[10.1073/pnas.0404206101](https://doi.org/10.1073/pnas.0404206101)
17. D.P. Martin, B. Murrell, M. Golden, A. Khoosal, B. Muhire, *Virus. Evol.* **1**, vev003 (2015)
18. G. Romay, H. Lecoq, C. Desbiez, *Arch. Virol.* **159**, 277–289 (2014). doi:[10.1007/s00705-013-1798-0](https://doi.org/10.1007/s00705-013-1798-0)
19. S. Yakoubi, H. Lecoq, C. Desbiez, *Virus. Genes.* **37**, 103–109 (2008). doi:[10.1007/s11262-008-0237-x](https://doi.org/10.1007/s11262-008-0237-x)