Arch Virol (2008) 153:1407–1410
DOI 10.1007/s00705-008-0123-9

ANNOTATED SEQUENCE RECORD

A new African streak virus species from Nigeria

Sunday Oluwafemi · Arvind Varsani · Adérito L. Monjane · Dionne N. Shepherd · Betty E. Owor · Edward P. Rybicki · Darren P. Martin

Received: 15 February 2008 / Accepted: 19 May 2008 / Published online: 3 June 2008

Abstract The African streak viruses (AfSVs) are a diverse group of mastrevirus species (family Geminiviridae) that infect a wide variety of annual and perennial grass species across the African continent and its nearby Indian Ocean islands. Six AfSV species (of which maize streak virus is the best known) have been described. Here we report the full genome sequences of eight isolates of a seventh AfSV species: Urochloa streak virus (USV), sampled from various locations in Nigeria. Despite there being good evidence of recombination in many other AfSV species, we found no convincing evidence that any of the USV sequences were either inter- or intra-species recombinants. The USV isolates, all of which appear to be variants of the same strain (their genome sequences are all more than 98% identical), share less than 69% nucleotide sequence identity with other currently described AfSV species.

Electronic supplementary material The online version of this article (doi:10.1007/s00705-008-0123-9) contains supplementary material, which is available to authorized users.

S. Oluwafemi
Department of Crop, Soil and Environmental Management, Bowen University, P.M.B 284, Iwo, Osun State, Nigeria

A. Varsani (✉)
Electron Microscope Unit, University of Cape Town, Private Bag, Rondebosch, Cape Town 7701, South Africa
e-mail: arvind.varsani@uct.ac.za

A. Varsani · A. L. Monjane · D. N. Shepherd · B. E. Owor · E. P. Rybicki
Department of Molecular and Cell Biology, University of Cape Town, Private Bag, Rondebosch 7701, South Africa

E. P. Rybicki · D. P. Martin
Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Observatory, Anzio Rd, Cape Town 7925, South Africa

The African streak viruses (AfSVs) are members of a diverse group of mastrevirus species (family *Geminiviridae*) that have been found infecting a wide variety of annual and perennial grass species across the African continent and its nearby Indian Ocean islands. Besides the best known AfSV, *Maize streak virus* (MSV), five other species have been described (those with italicised names are currently accepted by the ICTV): Eragrostis streak virus (ESV [20]); *Panicum streak virus* (PanSV [5]); *Sugarcane streak virus* (SSV [8]); *Sugarcane streak Egypt virus* (SSEV [2]) and *Sugarcane streak Reunion virus* (SSRV [17]). Here, we report the full genome sequences of eight isolates of a seventh proposed AfSV species: Urochloa streak virus (USV).

Eight *Urochloa deflexa* (common name: signal grass) plants displaying what appeared to be symptoms characteristic of PanSV infection (thin white discontinuous veinal streaks) were sampled from the following locations in Nigeria: Iwo (one plant, USV-[NIwo]: lat 7.62595°, lon 4.17803°), Ejigbo (two plants, USV-[NEji1]: lat 7.19889°, lon 4.32097°; USV-[NEji2]: lat 7.9122°, lon 4.31297°), Odo Oba (one plant, USV-[NOba]: lat 7.46381°, lon 4.13646°), Ile Igbo (one plant, USV-[NIle]; lat 7.61211°, lon 4.24097°), Ipetumodu (one plant, USV-[NIpe]: lat 7.51667°, lon 4.45000°) and Lagbaka (two plants, USV-[NLag1]: lat 8.91667°, lon 4.66667°; USV-[NLag2]: lat 8.92724°, lon 4.64886°). USV genomes were cloned from the eight samples following total DNA isolation by either CTAB- or Extract-*n*-Amp®-based protocols and amplification using

the ϕ29 DNA polymerase (TempliPhi^{TM}, GE Healthcare) methods described by Owor et al. [14, 15] and Shepherd et al. [19]. *Kpn*I- or *Bam*HI-digested full-length monomeric (2.7-kb) genomes were inserted into pGEM3Zf(+) (Promega Biotech) and sequenced by Macrogen Inc. (Korea) using primer walking.

We aligned the eight USV sequences with representative PanSV (PanSV-A[SKar]: L39638, PanSV-B[Ken]: X60168, PanSV-C[ZmGur]: EU224264, PanSV-D[NIfo]: EU224265); MSV (MSV-A[SMatA]: AF329881, MSV-D[SRaw]: AF329889, MSV-E[SPat]: AF329888, MSV-C[SSet]: AF007881, MSV-B[SVW]: AF239960); SSV (SSV-A[SN]: M82918, SSV-B[RPie]: EU244914); SSEV (SSVEV-Naga: AF239159); SSRV (SSRV-B[ZmNya]: EU244916; SSRV-A[Reu]: AF072672); ESV (ESV[ZmGur]: EU244915) and Digitaria streak virus (DSV: M230022) sequences using both ClustalW (gap open penalty = 10; gap extension penalty = 5; [23]) and manual adjustment of indel positions with the Mega 4.0 alignment editor [22].

We found evidence of various genomic features within the USV genomes by inference from those identified previously in other AfSV genomes. These included: (1) conserved inverted repeat sequences and iterated sequence elements on either side of a canonical geminiviral TAATATTAC sequence at the presumed virion-strand origin of replication (v-*ori*); (2) a conserved series of inverted repeat sequences immediately 3′ of the presumed complementary-strand origin of replication; (3) TATA boxes, GC-rich sequences and polyadenylation signals probably involved in complementary- and virion-sense gene transcription; (4) probable movement protein (*mp*), coat protein (*cp*), replication-associated protein (*rep*) and *repA* gene start and stop codons; and (5) probable *rep* and *mp* intron splice sites and branch point sequences (see Supplementary Fig. 1). The only peculiarity amongst these features was that all eight USV sequences had two distinct sets of iterated sequence elements that are potentially involved in v-*ori* recognition by Rep: all previously characterised AfSVs have only one set of these elements [1, 25]. By comparison with the predicted translation products of the various genes, we also identified (1) a conserved hydrophobic movement protein domain (Supplementary Fig. 2); (2) a potential nuclear localisation signal near the N-terminus of the coat protein (Supplementary Fig. 3); (3) conserved rolling-circle replication and retinoblastoma-related protein (pRBR) binding motifs in the replication-associated (Rep) and RepA proteins (Supplementary Figs. 4 and 5); and (4) potential *myb*-like transactivation domains and dNTP-binding motifs in the C-terminal half of the Rep protein (Supplementary Fig. 4).

Since inter-species recombination events have recently been detected amongst many of the AfSV sequences included in the alignment [20, 24], we analysed the full AfSV genome sequence alignment for evidence of recombination prior to phylogenetic analysis. We considered it possible that these recombination signals would complicate subsequent genetic distance and phylogenetic analyses aimed at relating the USV sequences to the other AfSV species. We therefore identified tracts of sequence that had a potentially recombinant origin using the RDP [9], GENECONV [16], BOOTSCAN [10], MAXCHI [13], CHIMAERA [10], SISCAN [6] and 3SEQ [3] methods implemented in RDP3 (default settings except that only recombination signals detectable by three or more different methods were considered [11]). Although we identified inter-species recombination signals previously detected in certain MSV, PanSV, SSEV and SSV isolates [12, 20, 24], we found no convincing evidence that any of the USV sequences were either inter- or intra-species recombinants (data not shown).

Following removal of tracts of sequence believed to have a recombinant origin from the alignment (sites 1,286–1,335 of MSV-B[SVW], 1,373–1,659 of ESV, 826–2,410 of SSEV-Naga, 961–1,040 of PanSV-D[NIfo], 336–550 and 2,215–2,256 of SSV-A[SN] and 338–548 and 2,210–2,258 of SSV-B[RPie]: all coordinates relative to the v-*ori* AC sequence), we used it with PHYML [7] to construct a bootstrapped (100 iterations) maximum-likelihood phylogenetic tree (Fig. 1). The evolutionary model used for the phylogeny reconstruction (F81 + G_4) was selected from amongst those available in PHYML using the same method as that employed by the MODELTEST web server [18]. We additionally calculated pair-wise sequence identities (with pair-wise discounting of sites with gaps) shared between the USV isolates and the other sequences in the alignment using RDP3.

It is clear from the phylogenetic and pairwise-identity analyses (Fig. 1) that the USV isolates belong to a new AfSV species. The sequences share <69% identity with isolates of the existing AfSV species (the ICTV guideline for mastrevirus species demarcation is <75% identity with a previously described species [21] and are all closely clustered on a distinct, well-supported branch of the phylogenetic tree. The placement of the USV branch within the tree is, however, reasonably uncertain. This is expressed in the presented maximum likelihood tree by the PanSV, SSV/SSEV/SSRV/ESV and USV lineages all branching from the same node. Whereas a similar trichotomy was encountered when another evolutionary model (HKY + G_4) was used to construct the maximum-likelihood tree, it was not seen in a neighbour-joining tree (K2P + G model, transition:transversion ratio = 2.0 in Mega 4.0; data not presented). In the neighbour-joining tree there was 92% bootstrap support for USV being slightly more closely related to viruses on the SSV/SSEV/SSRV/ESV branch than to those on the PanSV branch.

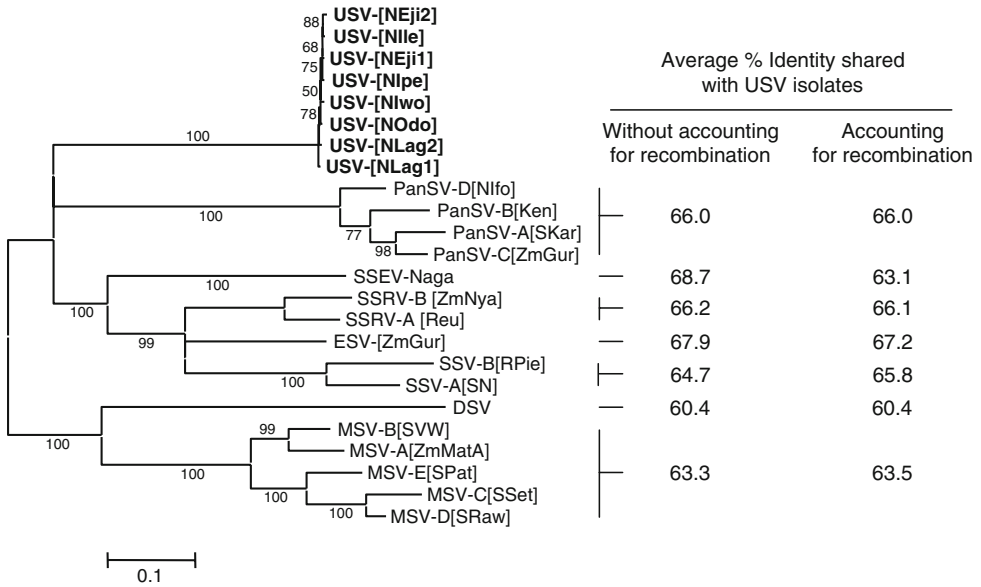

Fig. 1 Phylogenetic relationships amongst USV isolates and the other African streak viruses (*AfSVs*). This is an unrooted "recombination-free" maximum-likelihood tree ($F81 + G_4$) indicating the relationships between the USV sequences (in *bold*) and other representative AfSVs (Digitaria streak virus is not an AfSV but is included for comparative purposes). Numbers associated with *tree branches* indicate the percentage of 100 full maximum-likelihood bootstrap replicates supporting the existence of the branches. Branches with less than 50% bootstrap support have been collapsed. The mean pairwise sequence identities shared by the eight USV isolates and strains of the other AfSV species, both with and without taking recombination into account, are presented on the right-hand side of the tree. The *scale bar* represents nucleotides

Regardless of the phylogenetic uncertainty in the early origins of the USV lineage, the USV sequences all share >98.7% identity and are therefore all clearly variants of the same strain. This is perhaps not surprising, since they were all sampled from *U. deflexa* plants growing within 300 km of one another. This degree of diversity is very similar to that observed amongst PanSV strain A isolates sampled in South Africa, MSV strain A isolates sampled in Uganda [14] and SSEV isolates sampled in Egypt [2].

The distribution of USV throughout Africa remains to be determined. Whereas members of AfSV species such as MSV and PanSV have been detected throughout the continent [4, 12], [24], others such as SSEV (and now USV) have only ever been detected in individual countries. While these distribution differences may reflect sampling bias, it may also indicate that there are variable geographical and/or ecological barriers that different AfSV species experience during their movement around the continent.

Genbank accession numbers:

USV-NLag1: EU445692
USV-NIwo: EU445693
USV-NLag2: EU445694
USV-NOdo: EU445695
USV-NIle: EU445696
USV-NIpe: EU445697
USV-NEji: EU445698
USV-NEji2: EU445699

Acknowledgments This research was partially funded by the National Research Foundation (South Africa). A. V. is supported by the Carnegie Corporation of New York. D. N. S. is supported by PANNAR (Pty) Ltd; D. P. M. is supported by the Wellcome trust; B. O. is supported by the Rockefeller foundation through the USHEPiA programme; A. L. M. is supported by the Canon Collins Trust for Southern Africa and a University of Cape Town International Scholarship.

References

1. Argüello-Astorga GR, Guevara-Gonzalez RG, Herrera-Estrella LR, Rivera-Bustamante RF (1994) Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication. Virology 203:90–100
2. Bigarré L, Salah M, Granier M, Frutos R, Thouvenel J-C, Peterschmitt M (1999) Nucleotide sequence evidence for three distinct sugarcane streak mastreviruses. Arch Virol 144:2331–2344
3. Boni M, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176:1035–1047
4. Briddon RW, Lunness P, Chamberlin LC, Markham PG (1994) Analysis of the genetic variability of maize streak virus. Virus Genes 9:93–100
5. Briddon RW, Lunness P, Chamberlin LC, Pinner MS, Brundish H, Markham PG (1992) The nucleotide sequence of an infectious insect-transmissible clone of the geminivirus Panicum streak virus. J Gen Virol 73:1041–1047
6. Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16:573–582

7. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704
8. Hughes FL, Rybicki EP, von Wechmar MB (1992) Genome typing of southern African subgroup-1 geminiviruses. J Gen Virol 73:1031–1040
9. Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16:562–563
10. Martin DP, Posada D, Crandall KA, Williamson C (2005) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res Hum Retroviruses 21:98–102
11. Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21:260–262
12. Martin DP, Willment JA, Billharz R, Velders R, Odhiambo B, Njuguna J, James D, Rybicki EP (2001) Sequence diversity and virulence in *Zea mays* of *Maize streak virus* isolates. Virology 288:247–255
13. Maynard Smith J (1992) Analyzing the mosaic structure of genes. J Mol Evol 34:126–129
14. Owor BE, Martin DP, Shepherd DN, Edema R, Monjane AL, Rybicki EP, Thomson JA, Varsani A (2007) Genetic analysis of maize streak virus isolates from Uganda reveals widespread distribution of a recombinant variant. J Gen Virol 88:3154–3165
15. Owor BE, Shepherd DN, Taylor NJ, Edema R, Monjane AL, Thomson JA, Martin DP, Varsani A (2007) Successful application of FTA Classic Card technology and use of bacteriophage phi29 DNA polymerase for large-scale field sampling and cloning of complete maize streak virus genomes. J Virol Methods 140:100–105
16. Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology 265:218–225
17. Peterschmitt M, Reynaud B, Sommermeyer G, Baudin P (1991) Characterization of maize streak virus isolates using monoclonal and polyclonal antibodies and by transmission to a few hosts. Plant Dis 75:27–32
18. Posada D (2006) ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online. Nucleic Acids Res 34:700–703
19. Shepherd DN, Martin DP, Lefeuvre P, Monjane AL, Owor BE, Rybicki EP, Varsani A (2008) A protocol for the rapid isolation of full geminivirus genomes from dried plant tissue. J Virol Methods 149:97–102
20. Shepherd DN, Varsani A, Windram OP, Lefeuvre P, Monjane AL, Owor BE, Martin DP (2008) Novel sugarcane streak and sugarcane streak Reunion mastreviruses from southern Africa and La Reunion. Arch Virol 153:605–609
21. Stanley J, Bisaro DM, Briddon RW, Brown JK, Fauquet CM, Harrison BD, Rybicki EP, Stenger DC (2005) Geminiviridae. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (eds) Virus taxonomy (VIIIth report of the ICTV). Elsevier/Academic Press, London, pp 301–306
22. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599
23. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
24. Varsani A, Oluwafemi S, Windram OP, Shepherd DN, Monjane AL, Owor BE, Rybicki EP, Lefeuvre P, Martin DP (2008) Panicum streak virus diversity is similar to that observed for maize streak virus. Arch Virol 153:585–589
25. Willment JA, Martin DP, Palmer KE, Schnippenkoetter WH, Shepherd DN, Rybicki EP (2007) Identification of long intergenic region sequences involved in maize streak virus replication. J Gen Virol 88:1831–1841