Deeply sequenced infectious clones of key cassava begomovirus isolates from Cameroon

J. Steen Hoyer[1,#], Vincent N. Fondong[2], Mary M. Dallas[3], Catherine Doyle Aimone[3], David O. Deppong[3], Siobain Duffy[1,#], Linda Hanley-Bowdoin[3]

[1]Department of Ecology, Evolution, and Natural Resources, Rutgers University, New Brunswick, New Jersey, USA

[2]Department of Biological Sciences, Delaware State University, Dover, Delaware, USA

[3]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, North Carolina, USA

Running head: Full sequences for key clones of ACMV and EACMCV

Address correspondence to: J. Steen Hoyer, j.s.hoyer@wustl.edu, or Siobain Duffy, duffy@sebs.rutgers.edu

Abstract

We deeply sequenced two pairs of widely used infectious clones (4 plasmids) of the bipartite begomoviruses African cassava mosaic virus (ACMV) and East African cassava mosaic Cameroon virus (EACMCV). The sequences of the ACMV clones were quite divergent from our expectations. We have made raw reads, consensus plasmid sequences, and the infectious clones themselves publicly available.

Main text

Infectious clones are a central tool of molecular virology. Circular single-stranded DNA viruses such as begomoviruses are often cloned in a two-step process to create partial tandem dimers containing two copies of the virus origin-of-replication hairpin. This configuration is thought to enhance infection by facilitating release of monomer virus segment units by the viral Replication protein (Stenger et al., 1991). Cloned isolates of African cassava mosaic virus (ACMV) and East African cassava mosaic Cameroon virus (EACMCV) from Cameroon provided conclusive proof of synergy between two major clades of cassava begomoviruses (Fondong et al., 2000), which was a defining feature of an epidemic of mosaic disease that devastated cassava production in sub-Saharan Africa (Mbewe et al., 2020; Pita et al., 2001). These clones have been used extensively for molecular genetic analysis (Amin et al., 2011; Beyene et al., 2016; Chauhan et al., 2018; Chellappan et al., 2004, 2005a, 2005b; Chowda Reddy et al., 2008, 2009, 2012; Fondong et al., 2007; Kuria et al., 2017; Ndunguru et al., 2016; Patil et al., 2016; Pita et al., 2001; Reyes et al., 2013; Vanitharani et al., 2003, 2004), including analysis of spontaneous mutation (Aimone et al., 2020; Chen et al., 2019; Fondong and Chen, 2011). Herein we describe new sequence resources for these clones, which essentially confirm the sequence of the EACMCV clones and clarify the identity of the ACMV clones.

Complete and accurate plasmid sequences considerably simplify molecular analysis of infectious clones and design of new constructs. To confirm the sequences of the four plasmids listed in Table 1, we grew transformed *E. coli* DH5α cultures overnight at 37 ºC with ampicillin selection and purified each plasmid with the Qiagen Plasmid Maxi kit. Libraries were prepared from Covaris-sheared plasmid DNA in triplicate with the NEBNext Ultra II kit and sequenced on the Illumina NextSeq 500 platform in the 150 bp paired-end read configuration. Data are available as SRA BioProject PRJNA649777.

Three of these plasmids were described by Fondong et al. (2000), and the fourth, for EACMCV DNA-B, was described later (Chowda Reddy et al., 2012; Fondong and Chen, 2011) due to greater difficulty in cloning that sequence in the partial-tandem-dimer configuration. These plasmids had not previously been fully sequenced, so we deduced sequence maps based on the restriction sites used, including for partial-tandem-dimer virus segment inserts. Reads were trimmed with CutAdapt 1.16 (Martin, 2011) and aligned to the hypothesized sequences with the Burrows-Wheeler Aligner (BWA-MEM v0.7.13 (Li, 2013)). Variants relative to each reference sequence were

identified with samtools v1.8 (Li et al., 2009) and VarScan v2.4.4 (Koboldt et al., 2012). We corrected each plasmid sequence (Table 1) and aligned reads to it a second time.

Table 1. GenBank and Addgene identifiers for the four previously described infectious clones and corresponding virus (monomer) sequences. Six complete sequences are described for the first time here, whereas AF112354 and FJ826890 were previously described (Chowda Reddy et al., 2012; Fondong and Chen, 2011; Fondong et al., 2000).

	GenBank accession		
	Plasmid	Virus	Addgene ID
pBluescript II KS(+) **ACMV** DNA-**A** 1.4mer	MT856193	MT858793	159134
pUC19 **ACMV** DNA-**B** 1.5mer	MT856194	MT858794	159135
pBluescript II KS(+) **EACMCV** DNA-**A** 1.6mer	MT856195	AF112354	159136
pSL1180 **EACMCV** DNA-**B** 1.6mer	MT856192	FJ826890	159137

The EACMCV DNA-A and DNA-B clones had four and two single-nucleotide differences relative to their corresponding GenBank accessions (AF112354.1 and FJ826890.1). We list these differences in the standard coordinate system (relative to the virus nick site) in the order new sequence-old sequence: T139A, G161R [note ambiguity code in AF112354.1], T181TC, and A206AC for DNA-A and T1671G, A2724AT for DNA-B. These differences may reflect errors or they may be true differences resulting from mutations that occurred in *E. coli*.

The consensus sequences of the ACMV clones, by contrast, were 3.1% and 5.8% divergent from the sequences (AF112352.1 and AF112353.1) originally reported by Fondong et al. (2000), as calculated with Sequence Demarcation Tool v1.2 (Muhire et al., 2014). This difference was not entirely unexpected, because of the parallel history of two sets of ACMV clones: infectious partial-tandem-dimer clones were made via restriction-digestion/ligation from sap-inoculated *Nicotiana benthamiana* plants whereas the monomer-segment-unit clones were cloned with PCR from the same original cassava field sample (Fondong et al., 2000). We expect these complete infectious clone sequences will be of great utility to the community, given that many follow-up publications (Amin et al., 2011; Beyene et al., 2016; Chauhan et al., 2018; Chellappan et al., 2004, 2005a, 2005b; Kuria et al., 2017; Ndunguru et al., 2016; Patil et al., 2016; Reyes et al., 2013; Vanitharani et al., 2003, 2004) specifically referenced the

related but non-identical monomer sequences (AF112352.1 and AF112353.1).

We obtained deep coverage, over 18,000-fold across all positions for all four plasmids with an average of 157,000-fold coverage (Figure 1). This read depth was consistent across three separate libraries for each plasmid (see Zenodo record 4075362).

Our results underscore the value of confirming the sequence of molecular clones. An similar example from the geminivirus literature is the Nigerian Ogoroco clone of ACMV (Briddon et al., 1998). Kittelmann et al. (Kittelmann et al., 2009) thoroughly documented that the infectious DNA-A clone is not 100% identical to the related GenBank accession (AJ427910.1). We do not believe the large literature on these Cameroonian ACMV isolates requires dramatic reinterpretation, but we recommend using the revised sequences described here for future analyses.

Data availability: Plasmids are available from Addgene and sequences for full plasmids and ACMV segments are available in GenBank (Table 1). Raw Illumina data are available from the NCBI Sequence Read Archive (PRJNA649777). Data processing code has been archived as Zenodo record 4075362.

Acknowledgements: This work was supported by NSF award OIA-1545553 to LHB and SD. We thank the NC State University Genomic Sciences Laboratory (Raleigh, NC, USA) for excellent sequencing service. We thank the staff of the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey, for access to and maintenance of the Amarel cluster. We thank Getu Beyene (Donald Danforth Plant Science Center) for helpful discussions.

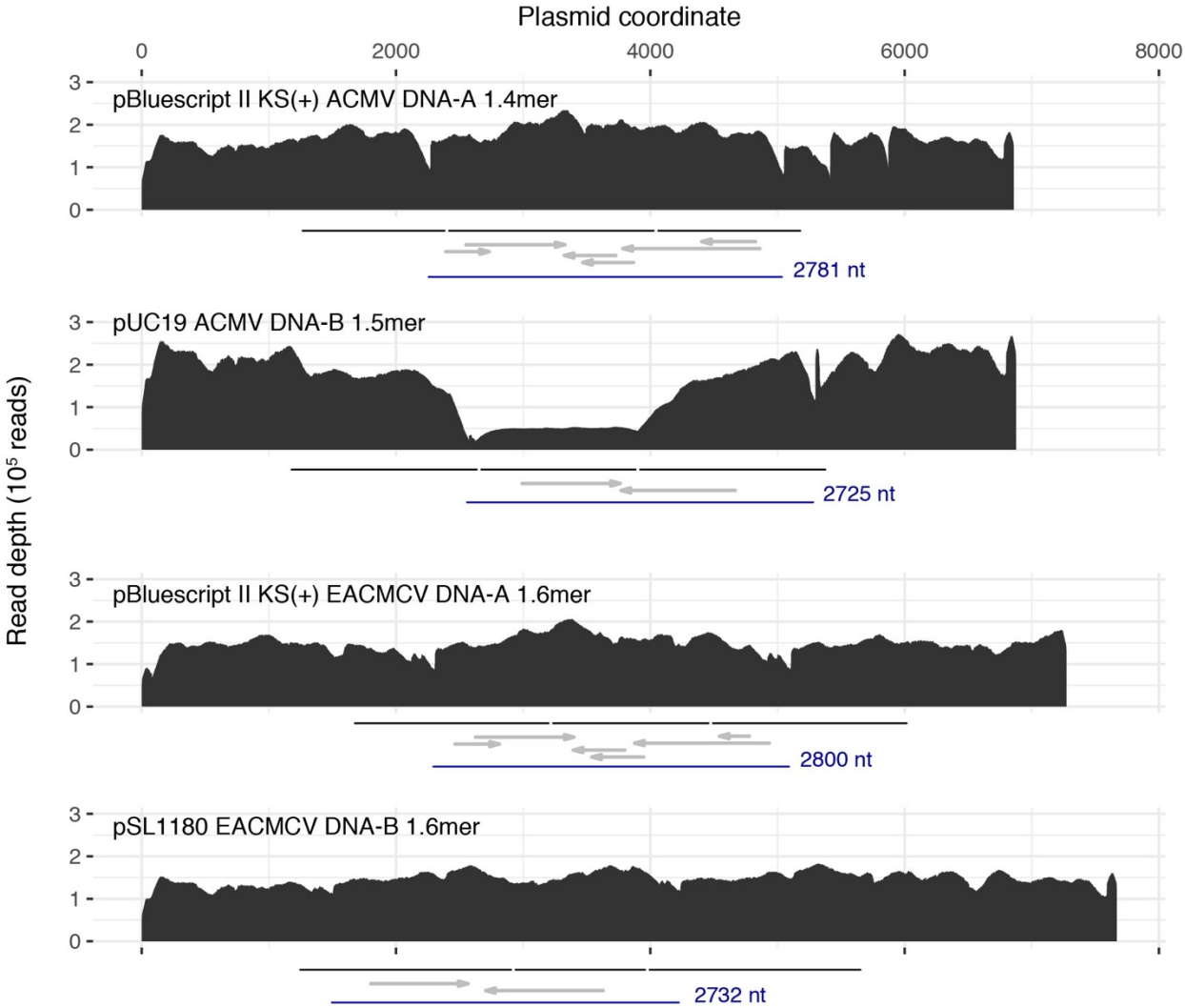

Figure 1. Plots of Illumina read depth across the length of the four infectious clone plasmids. One of three libraries is shown for each plasmid (SRR12354432, SRR12354427, SRR12354424, and SRR12354421). The region in each plasmid corresponding to each virus segment partial-tandem-dimer unit is indicated with a black line under each graph. Vertical white lines demarcate the boundaries of the unique and duplicated regions of each concatemer. Each virus segment monomer unit (in between two replication-origin nick sites) is shown in blue. Canonical virus genes are indicated with gray arrows, left to right for virus sense (AV1, AV2, BV1) and right to left for complementary sense (AC1 to AC4, BC1). Uneven read depth for the ACMV DNA-B plasmid is due to instability (truncation), which is evident in single-cut restriction digests (not shown). Such partial deletion of tandem duplicated regions in *E. coli* is not uncommon (Oliveira et al., 2009).

References

Aimone, C.D., Hoyer, J.S., Dye, A.E., Deppong, D.O., Duffy, S., Carbone, I., and Hanley-Bowdoin, L. (2020). An improved experimental pipeline for preparing circular ssDNA viruses for next-generation sequencing. BioRxiv 2020.09.30.321224. https://doi.org/10.1101/2020.09.30.321224

Amin, I., Patil, B.L., Briddon, R.W., Mansoor, S., and Fauquet, C.M. (2011). A common set of developmental miRNAs are upregulated in Nicotiana benthamiana by diverse begomoviruses. Virology Journal *8*, 143. https://doi.org/10.1186/1743-422X-8-143

Beyene, G., Chauhan, R.D., Wagaba, H., Moll, T., Alicai, T., Miano, D., Carrington, J.C., and Taylor, N.J. (2016). Loss of CMD2-mediated resistance to cassava mosaic disease in plants regenerated through somatic embryogenesis. Molecular Plant Pathology *17*, 1095–1110. https://doi.org/10.1111/mpp.12353

Briddon, R.W., Liu, S., Pinner, M.S., and Markham, P.G. (1998). Infectivity of African cassava mosaic virus clones to cassava by biolistic inoculation. Arch. Virol. *143*, 2487–2492. https://doi.org/10.1007/s007050050478

Chauhan, R.D., Beyene, G., and Taylor, N.J. (2018). Multiple morphogenic culture systems cause loss of resistance to cassava mosaic disease. BMC Plant Biology *18*, 132. https://doi.org/10.1186/s12870-018-1354-x

Chellappan, P., Vanitharani, R., and Fauquet, C.M. (2004). Short interfering RNA accumulation correlates with host recovery in DNA virus-infected hosts, and gene silencing targets specific viral sequences. J. Virol. *78*, 7465–7477. https://doi.org/10.1128/JVI.78.14.7465-7477.2004

Chellappan, P., Vanitharani, R., Ogbe, F., and Fauquet, C.M. (2005a). Effect of temperature on geminivirus-induced RNA silencing in plants. Plant Physiol. *138*, 1828–1841. https://doi.org/10.1104/pp.105.066563

Chellappan, P., Vanitharani, R., and Fauquet, C.M. (2005b). MicroRNA-binding viral protein interferes with Arabidopsis development. PNAS *102*, 10381–10386. https://doi.org/10.1073/pnas.0504439102

Chen, K., Khatabi, B., and Fondong, V.N. (2019). The AC4 protein of a cassava geminivirus is required for virus infection. MPMI. https://doi.org/10.1094/MPMI-12-18-0354-R

Chowda Reddy, R.V., Achenjang, F., Felton, C., Etarock, M.T., Anangfac, M.-T., Nugent, P., and Fondong, V.N. (2008). Role of a geminivirus AV2 protein putative protein kinase C motif on subcellular localization and pathogenicity. Virus Research *135*, 115–124. https://doi.org/10.1016/j.virusres.2008.02.014

Chowda Reddy, R.V., Dong, W., Felton, C., Ryman, D., Ballard, K., and Fondong, V.N. (2009). Characterization of the cassava geminivirus transcription activation protein putative nuclear localization signal. Virus Research *145*, 270–278. https://doi.org/10.1016/j.virusres.2009.07.022

Chowda Reddy, R.V., Dong, W., Njock, T., Rey, M.E.C., and Fondong, V.N. (2012). Molecular interaction between two cassava geminiviruses exhibiting cross-protection. Virus Research *163*, 169–177. https://doi.org/10.1016/j.virusres.2011.09.009

Fondong, V.N., and Chen, K. (2011). Genetic variability of East African cassava mosaic Cameroon virus under field and controlled environment conditions. Virology *413*, 275–282. https://doi.org/10.1016/j.virol.2011.02.024

Fondong, V.N., Pita, J.S., Rey, M.E.C., de Kochko, A., Beachy, R.N., and Fauquet, C.M. (2000). Evidence of synergism between African cassava mosaic virus and a new double-recombinant geminivirus infecting cassava in Cameroon. Journal of General Virology *81*, 287–297. https://doi.org/10.1099/0022-1317-81-1-287

Fondong, V.N., Reddy, R.V.C., Lu, C., Hankoua, B., Felton, C., Czymmek, K., and Achenjang, F. (2007). The Consensus N-Myristoylation Motif of a Geminivirus AC4 Protein Is Required for Membrane Binding and Pathogenicity. MPMI *20*, 380–391. https://doi.org/10.1094/MPMI-20-4-0380

Kittelmann, K., Rau, P., Gronenborn, B., and Jeske, H. (2009). Plant geminivirus Rep protein induces rereplication in fission yeast. Journal of Virology *83*, 6769–6778. https://doi.org/10.1128/JVI.02491-08

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568–576. https://doi.org/10.1101/gr.129684.111

Kuria, P., Ilyas, M., Ateka, E., Miano, D., Onguso, J., Carrington, J.C., and Taylor, N.J. (2017). Differential response of cassava genotypes to infection by cassava mosaic geminiviruses. Virus Research *227*, 69–81. https://doi.org/10.1016/j.virusres.2016.09.022

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 [q-Bio]. http://arxiv.org/abs/1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12. https://doi.org/10.14806/ej.17.1.200

Mbewe, W., Hanley-Bowdoin, L., Ndunguru, J., and Duffy, S. (2020). Cassava viruses: Epidemiology, evolution, and management. In Emerging Plant Diseases and Global Food Security, J.B. Ristaino, and A. Records, eds. (The American Phytopathological Society), pp. 133–157. https://doi.org/10.1094/9780890546383.007

Muhire, B.M., Varsani, A., and Martin, D.P. (2014). SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. PLOS ONE *9*, e108277. https://doi.org/10.1371/journal.pone.0108277

Ndunguru, J., León, L.D., Doyle, C.D., Sseruwagi, P., Plata, G., Legg, J.P., Thompson, G., Tohme, J., Aveling, T., Ascencio-Ibáñez, J.T., and Hanley-Bowdoin, L. (2016). Two novel DNAs that enhance symptoms and overcome CMD2 resistance to cassava mosaic disease. Journal of Virology *90*, 4160–4173. https://doi.org/10.1128/JVI.02834-15

Oliveira, P.H., Prather, K.J., Prazeres, D.M.F., and Monteiro, G.A. (2009). Structural instability of plasmid biopharmaceuticals: challenges and implications. Trends in Biotechnology *27*, 503–511. https://doi.org/10.1016/j.tibtech.2009.06.004

Patil, B.L., Bagewadi, B., Yadav, J.S., and Fauquet, C.M. (2016). Mapping and identification of cassava mosaic geminivirus DNA-A and DNA-B genome sequences for efficient siRNA expression and RNAi based virus resistance by transient agro-infiltration studies. Virus Research *213*, 109–115. https://doi.org/10.1016/j.virusres.2015.11.011

Pita, J.S., Fondong, V.N., Sangaré, A., Otim-Nape, G.W., Ogwal, S., and Fauquet, C.M. (2001). Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. Journal of General Virology *82*, 655–665. https://doi.org/10.1099/0022-1317-82-3-655

Reyes, M.I., Nash, T.E., Dallas, M.M., Ascencio-Ibáñez, J.T., and Hanley-Bowdoin, L. (2013). Peptide aptamers that bind to geminivirus replication proteins confer a resistance phenotype to tomato yellow leaf curl virus and tomato mottle virus infection in tomato. Journal of Virology *87*, 9691–9706. https://doi.org/10.1128/JVI.01095-13

Stenger, D.C., Revington, G.N., Stevenson, M.C., and Bisaro, D.M. (1991). Replicational release of geminivirus genomes from tandemly repeated copies: evidence for rolling-circle replication of a plant viral DNA. PNAS *88*, 8029–8033. https://doi.org/10.1073/pnas.88.18.8029

Vanitharani, R., Chellappan, P., and Fauquet, C.M. (2003). Short interfering RNA-mediated interference of gene expression and viral DNA accumulation in cultured plant cells. PNAS *100*, 9632–9636. https://doi.org/10.1073/pnas.1733874100

Vanitharani, R., Chellappan, P., Pita, J.S., and Fauquet, C.M. (2004). Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. J. Virol. *78*, 9487–9498. https://doi.org/10.1128/JVI.78.17.9487-9498.2004