

2017 International Symposium on Cocoa Research (ISCR), Lima, Peru, 13-17 November 2017

Molecular genomic diversity of previously undescribed cacao swollen shoot badnaviruses in Nigeria

L. Dongo¹, N. Chingandu² and J.K. Brown²

¹Cocoa Research Institute of Nigeria, Ibadan, Nigeria; ²School of Plant Sciences, The University of Arizona, Tucson, Arizona 85721, USA

Abstract

Cacao swollen shoot disease (CSSD) is caused by a complex of badnaviruses and is a major pathogen infecting cocoa in West Africa, causing as much as 100% yield loss. The shortage of knowledge about these viruses, their distribution, and genomic variability has precluded the development of molecular diagnostic tools, essential for breeding programs aimed at developing cocoa tolerant or resistant genotypes. To characterize the genomic variation of CSSD badnavirus isolates in Nigeria, forty-nine cocoa leaf samples exhibiting virus-like symptoms were collected from the Oyo, Osun and Ondo states. CSSD-badnavirus presence was confirmed in 32 of 49 leaf samples (65.3%) by PCR amplification of a 577 bp region of the intergenic region (IR). The phylogenetic analysis resolved four clades, two well-supported and two unsupported major clades. About 29% were found to be closely related to those previously sequenced from Ghana and Ivory Coast (GenBank accessions), but none were related to the three published Togo isolate sequences. Two thirds of the isolates (60 of 84; 71%) from Nigeria grouped into a single clade composed of two groups or strains of a predicted, new CSSD-badnavirus species, thus far uniquely found in Nigeria. For all isolates combined, there was no apparent relationship between genotype and symptoms or geographical location of collection site, indicating genomic variability is distributed in cacao at the collection sites studied. The isolates that grouped with previously described CSSD badnaviruses are possibly the result of introductions of viral genotypes because of germplasm exchange among the three countries (Nigeria, Ghana and Cote d'Ivoire) during regional international germplasm projects. The full-length genome sequence was determined for selected isolates within the unique clade, revealing a previously unidentified species. The new species shared 70-75% nucleotide identity with other known CSSD badnaviruses, based on pairwise distance analysis of the taxonomically informative viral RT-RNase H region. The inability to PCR-amplify isolates from all symptomatic samples suggests the possible association of additional, undiscovered CSSD badnavirus-like isolates and/or other non-badnaviruses with symptomatic trees in Nigeria, highlighting the need for extensive studies to understand the full extent of CSSD badnavirus variability and other virus-like pathogens. This is the first report of a previously undescribed badnaviral genome associated with CSSD symptoms in cacao in Nigeria, herein designated, Cacao red vein-banding virus.

Keywords: cacao badnavirus, Cacao red vein-banding virus, mealybug transmitted virus, cacao swollen shoot virus Nigeria

Corresponding author: Lelia Dongo Lelia Dongo <leliadongo@yahoo.co.uk>

Introduction

Cocoa is the major foreign exchange earner for Nigeria, aside from crude oil. It is an important crop for development and poverty alleviation, providing a means of livelihood for approximately 450,000 farming families in Nigeria. Despite the efforts currently made by the Nigerian government to increase its share in the world market through intensification of cocoa production by rehabilitating old plantations and free distribution of improved planting materials, the country's production is on the decrease. Nigeria's cocoa production has dropped over the years and recently by 5.2 per cent from 248,000 metric tonnes in the 2013/2014 planting season to 235,000 metric tonnes in the 2014/2015 season. This decline has been attributed to several factors including pest and diseases. Cocoa pests and diseases are a major constraint to sustainable cocoa production as they account for about 40% of global loss of cocoa production annually. 15% of an estimated 47% of world cacao output lost due to pests and diseases has been attributed to cacao swollen shoot disease (CSSD) (COPAL, 2009).

To date, a number of badnaviruses have been recognized as causal pathogens of CSSD (Abrokwah *et al*; 2016, Chingandu *et al*; 2017a, Muller *et al*, 2018). They are considered endemic to West Africa and belong

to the genus *Badnavirus*, family, *Caulimoviridae*. CSSD causes significant losses to the cacao crop on an annual basis, and yield reduction can be as great as 100%. Infected trees may not show symptoms until one year or more after inoculation (presumed, infection), followed by decline and death in 3-5 years. The first symptoms of CSSD in Nigeria were reported in 1944 near the village of Egbeda, 13 miles east of Ibadan in Oyo State. Shortly afterwards, outbreaks became widespread, suggesting that the swollen shoot disease of cacao had occurred in Nigeria long before the first report (Lister and Thresh, 1958).

To manage the disease, during the 1950's, the government of Nigeria introduced a 'cutting out policy' to remove infected trees, thereby reducing sources of inoculum and preventing further spread. This resulted in the elimination of approximately 1.8 million trees between 1946 and 1956. However, this management strategy was abandoned because of opposition by the farmers when they were not compensated for the costs of tree removal and replanting. Consequently, the policy was replaced by the isolation of infected zones through cordon sanitaire. However, cut-out management practices had been applied to only 5% of the area known to be affected by the disease before the policy was abandoned, and presently, CSSD symptoms have been documented in five cocoa producing states (Edo, Cross river, Oyo, Ondo and Taraba) of Nigeria (Dongo and Orisajo, 2007; Dongo *et al.*, 2012).

Disease control after the failed cutting-out policy was attempted through the establishment of a cacao breeding program, with cacao introductions from the Upper Amazon (Posnette, 1951) but these varieties were of no immediate practical value due to the very low level of resistance or tolerance (Posnette and Todd, 1951). Management of CSSD is difficult due to a number of complexities, including lack of knowledge about the putative badnaviral causal agent(s), cacao host genotypic variability, the mealybug vector and its role (as many as 12 species) in relation to CSSD epidemiology, and socio-economic factors. And, virus detection in relation to time of infection and symptom development has not been fully resolved.

Breeding for disease resistance will require an understanding of the identity and distribution of the different species and strains of the CSSD badnavirus complex, and the development of reliable diagnostic tools to facilitate the selection and breeding of tolerant to resistant varieties. The genomic variation of badnaviruses associated with CSSD-like symptoms in West Africa has only been investigated in detail in very recent years, and evidence suggests that a complex of viruses is involved, and that new variants or species may have emerged somewhat recently.

This study was carried out to determine the extent of sequence variability among and between CSSD badnavirus isolates in Nigeria and to characterize any new cacao-infecting badnavirus species associated with CSSV.

Materials and Methods

To determine the extent of sequence variation between CSSD isolates in Nigeria, forty-nine leaf samples exhibiting CSSD-like symptoms were collected from Ondo, Osun, and Oyo states. Total nucleic acids were purified from leaves using the CTAB protocol of Doyle and Doyle (1990). Polymerase chain reaction (PCR) amplification was carried out using CSSVUni primer pairs designed around the seven reference CSSD viral genomes available in GenBank (unpublished data). Amplicons of the expected sizes were ligated into pGEM-T Easy vector, transformed into *E. coli* DH5 alpha cells, and sequenced bidirectionally using M13 universal primers. Based on the 577 base pair (bp) fragment of the intergenic region (IR), representatives of the most divergent isolates were selected for complete genome sequencing. To determine the complete genome sequence for the most divergent isolates in the Oyo and Osun states, eighteen additional samples exhibiting a range of diverse CSSD-like symptoms, including fern pattern, chlorotic mottling, vein clearing, necrotic crinkling, chlorosis, and red vein-banding, were collected. Based on an alignment of the IR sequences, abutting PCR primers were designed and used to amplify a full-length genome sequence, ~7 kbp in size from each DNA sample. The apparently full-length genomes were cloned and the DNA sequence was determined by primer walking, with at least a ~150 bp overlap.

The assembled sequences were provisionally identified using BLASTn (NCBI), and the coding regions were predicted using ORF Finder (NCBI). The sequences were aligned using MUSCLE (Edgar, 2004) implemented in CLC Sequence viewer 7, and analyzed phylogenetically using Maximum likelihood in MEGA6 (Tamura *et al.*, 2013). Pairwise distance estimates of the IR locus (577 bp), and taxonomic relationships based on the 1230 bp RT-RNase H locus of the full-length genomes, at 1230 bp, were carried out using the Standard Demarcation Tool (Muhire *et al.*, 2014).

Results and Discussion

Badnavirus isolates most closely related to CSSD isolates were detected in 33 of 49 samples (65.0%) by PCR amplification of the 577 bp non-coding IR, whereas, no badnaviruses were detected in the remaining 16 samples (Fig. 1). Collectively, the shared nucleotide (nt) identities among all 54 field and Genbank isolates ranged from 52-100%, suggesting the presence of both known and previously undiscovered badnaviruses associated with CSSD symptoms in Nigeria.

The BLASTn analysis (GenBank database) of the nine apparently complete badnaviral genomes obtained by PCR and sequenced by primer walking, indicated that their closest relatives were other CSSD-associated badnaviruses.

Pairwise nucleotide (distance) analysis of the taxonomically-relevant RT-RNase H sequence region, indicated that the nine isolates shared 70-75% nt identity at the RT-RNase H region with previously identified CSSD badnaviruses, whereas, the within group divergence was 86-99% (Fig. 2). Collectively, these results indicate the nine isolates showed sufficient divergence to constitute their recognition as a previously unidentified member of the CSSD badnavirus complex. The full-length genome sequences of this proposed new species were found to encode four predicted open reading frames, and the predicted, conserved protein domains were similar to those of other badnaviruses, including aspartate pepsin protease, reverse transcriptase, ribonuclease H, and zinc knuckle protein (Fig. 3).

Based on phylogenetic analysis (Fig.4) the nine isolates grouped in the same clade, separately from the four other CSSD-associated species, known thus far (see references in Chingandu et al., 2017a). Consequently, the combined results of the phylogenetic and distance analyses corroborated the extreme divergence of these nine Nigerian isolates, in relation to other previously known CSSD-associated badnaviruses, and support their recognition as a new member of the CSSD complex, herein designated Cacao red vein-banding virus (CRVBV).

Research Outlook

This discovery of another new badnavirus species associated with CSSD in West Africa highlights the need for additional research with respect to genome variability and functional characterization, evolutionary relationships between all species and strains thereof, epidemiological studies, toward the collective management of swollen shoot disease of cacao in West Africa. These newly determined genome sequences will further inform the design and development of molecular diagnostic assays to achieve sensitive and reliable CSSD-virus detection, in advance of and post-symptom development.

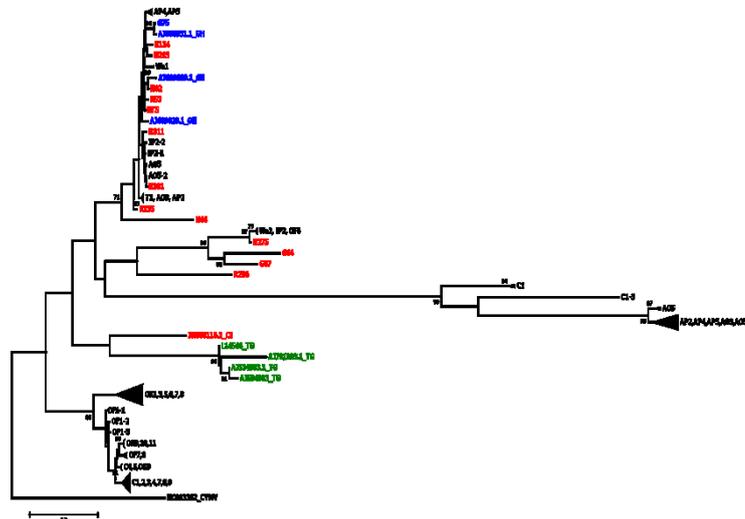


Fig. 1. Maximum likelihood tree ($\geq 70\%$, 1000 iterations) reconstructed for the intergenic region sequences of the seven original Genbank CSSD-badnavirus reference sequences, the 14 recently published and the 49 IR sequences for isolates from Cote d'Ivoire (CI); Ghana (GH), and Togo (TG). The Nigerian isolates are indicated with AP, AO, C, IP, OF, OK, T and WA codes.

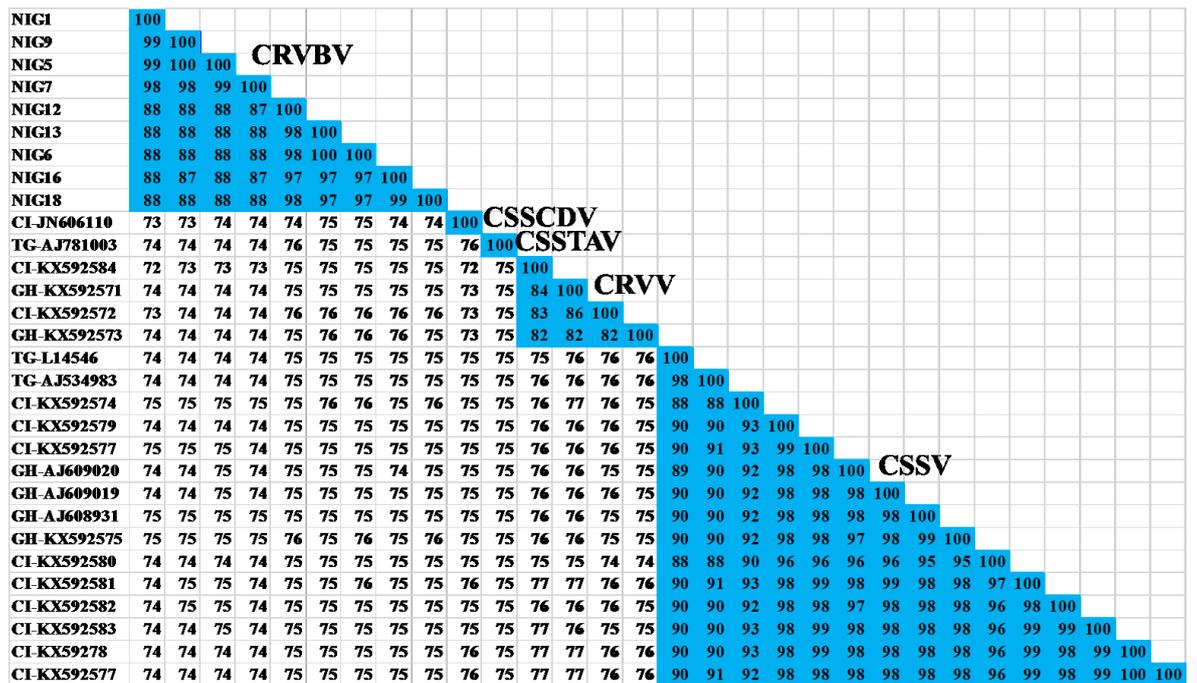
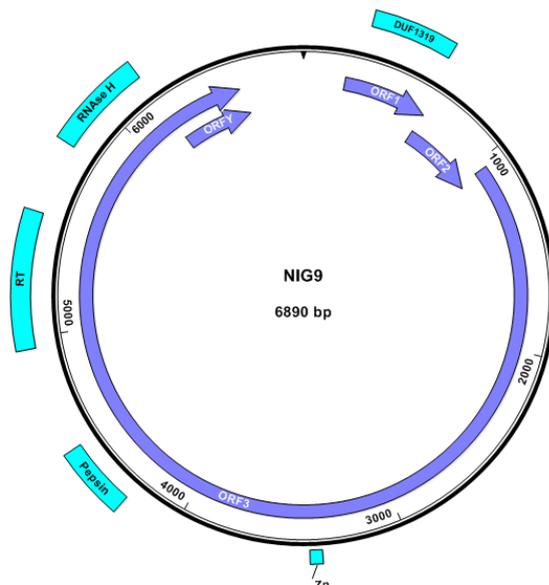


Fig. 2. Pairwise distance matrix for the taxonomically informative, RT-RNase H locus comprising 1,230 bases, using the Sequence Demarcation Tool v1.2 to predict relationships between the nine genome sequences determined for cacao isolates from Nigeria and other cacao-associated badnaviruses from West Africa. The highlighted cells represent $\geq 80\%$ shared nucleotide identity, which is considered a species. The nine Nigerian isolates belong to one species, Cacao red vein-banding virus (CRVBV), at 88 – 100% shared nucleotide identity. CRVBV is distinct from the four badnavirus species associated with cacao: Cacao red vein virus (CRVV), *Cacao swollen shoot CD virus* (CSSCDV), *Cacao swollen shoot Togo A virus* (CSSTAV), and *Cacao swollen shoot virus* (CSSV).



- 9 genomes sequenced
- Length 6820 – 6895 bp, shorter than all cacao badnavirus genomes: 7004 – 7211 bp
- All have 4 ORFs
- Similar conserved protein domains to CSSV

Code	Domain
DUF1319	Domain of unknown function
Zn	Zinc knuckle, coat protein motif
Pepsin	Pepsin-like aspartate protease
RT	Reverse transcriptase
RNase H	Ribonuclease H

Fig. 3. Genome arrangement and conserved protein domains of a representative of the nine isolates of CRVBV as predicted by NCBI open reading frame (ORF) Finder and Conserved domain database search. Filled arrows represent ORFs and boxes indicate the conserved protein domains.

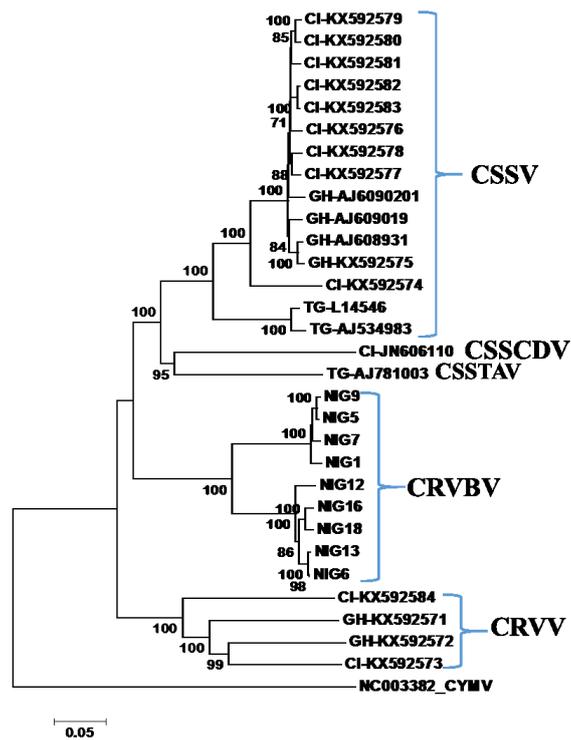


Fig. 4. Full-length genome phylogenetic analysis of nine sequences from Nigeria and other West-African cacao-infecting badnavirus species using Neighbor joining 1000 bootstrap replications implemented in MEGA6. CSSV – *Cacao swollen shoot virus*, CSSTAV – *Cacao swollen shoot Togo A virus*, CSSCDV – *Cacao swollen shoot CD virus*, CRVV – *Cacao red vein virus* (proposed), CYMV – *Citrus yellow mosaic virus*, CI– Cote d’Ivoire, GH – Ghana, NG – Nigeria, and TG – Togo.

References

- Abrokwha, F., Dzahini-Obiatay, H., Galyuon, I., Ossae Awuku, F. and Muller, E. (2016). Geographical distribution of cacao swollen shoot virus molecular variability in Ghana. *Plant Dis.* 100: 2011-2017
- Chingandu, N., Kouakou, K., Aka, R., Ameyaw, G., Gutierrez, O.A., and Brown, J.K. 2017a. The proposed new species, *Cacao red vein virus*, and three previously recognized badnaviruses are associated with cacao swollen shoot disease. *Viol. J.* 14:199, DOI 10.1186/s12985-017-0866-6.
- Chingandu, N., Kouakou, K., Aka, R., Gutierrez, O. A., and Brown, J. K. 2017b. Unexpected genome variability at multiple loci suggests *Cacao swollen shoot virus* comprises multiple, divergent molecular variants. *J Emerg Virol Dis* 3(1): doi <http://dx.doi.org/10.16966/2473-1846.128>.
- Dongo L. N. and Orisajo, S. B. (2007). Status of Cocoa Swollen Shoot Virus Disease in Nigeria. *African J. Biotechnol.* 6(17): 2054–2061.
- Dongo, L. N., Ohunakin, A. O., Atiri, G. I. and Kumar, P. L. (2012). Assessment of the Impact of Cacao Swollen Shoot Virus in Nigeria and Development of Simple Diagnostic Tools. In: Proceedings of the 17th International Cocoa Research Conference, Yaounde, 2012.
- Doyle, J. J., and Doyle, J. L. 1990. Isolation of plant DNA from fresh tissue. *Focus (Madison).* 12:13–15.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Lister, R.M. and Thresh, J.M. (1958). The history and control of cocoa swollen shoot disease in Nigeria. In *Cocoa, chocolate and Confectionery alliance Ltd. Report of the Cocoa Conference, 1957.* London. pp. 132-142.
- Muhire B, Varsani A, Martin DP (2014) SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 9:e108277
- Muller, E., Ravel, S., Agret, C., Abrokwha, F., Dzahini-Obiatay, H., Galyuon, I., et al. (2018). Next generation sequencing elucidates cacao badnavirus diversity and reveals the existence of more than ten viral species. *Virus Res.* 244:235–251.
- Posnette, A.F. and Todd, J.M. (1951). Virus disease of cocoa in West Africa VII. The search for virus resistant cocoa. *Ann. Appl. Biol.* 38: 785-800.
- Posnette, A.F. (1951). Virus research at the West African Cacao Research Institute, Tafo, Gold Coast. Govt Printing Office Trinidad. BWI-1952.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30:2725–2729.