

Molecular variability analysis of five new complete cacao swollen shoot virus genomic sequences

E. Muller¹ and S. Sackey²

¹CIRAD, UMR BGPI, Campus International de Baillarguet,
Montpellier, France

²University of Ghana, Department of Biochemistry,
Legon, Ghana

Received March 29, 2004; accepted July 12, 2004
Published online August 30, 2004 © Springer-Verlag 2004

Summary. Cacao swollen shoot virus (CSSV), a member of the family *Caulimoviridae*, genus *Badnavirus* occurs in all the main cacao-growing areas of West Africa. We amplified, cloned and sequenced complete genomes of five new isolates, two originating from Togo and three originating from Ghana. The genome of these five newly sequenced isolates all contain the five putative open reading frames I, II, III, X and Y described for the first sequenced CSSV isolate, Agou1 originating from Togo. Their genomes have been aligned with the genome of Agou1. The nucleotide and amino acid sequence identities between isolates have been calculated and a phylogenetic analysis has been made including other pararetroviruses. Maximum nucleotide sequence variability between complete genomes of CSSV isolates was 29.4%. Geographical differentiation between isolates appears more important than differentiation between mild and severe isolates. ORF X differs greatly in size and sequence between the Togolese isolates Nyongbo2 and Agou1, and the four other isolates, its functional role is therefore clearly questionable.

Introduction

Isolates of the species *Cacao swollen shoot virus* (CSSV) are mealybug-transmitted viruses occurring in all the main cacao-growing areas of West Africa. The characteristic disease symptoms are red vein banding in young unhardened leaves, and shoot, stem and root swelling. The most virulent isolates Agou1 (Togo) and New Juaben (Ghana) cause defoliation, small pods and death of the tree. Some mild isolates are associated with red vein banding and leaf chlorosis. CSSV, a member of the family *Caulimoviridae*, genus *Badnavirus*, possesses non-enveloped bacilliform particles and a double stranded DNA genome of 7.16 kb [17, 18, 9]. Five

putative open reading frames (ORFs) are located on the plus strand. ORF1 encodes a 16.7 kDa protein whose function is not yet determined. The ORF2 product is a 14.4 kDa nucleic acid-binding protein [16]. ORF3 codes for a polyprotein of 211 kDa which contains, from its amino- to carboxyl-terminus, consensus sequences for a cell-to-cell movement protein [9], an RNA binding domain of the coat protein, an aspartyl proteinase, a reverse transcriptase (RT), and a ribonuclease H (RNase H). The last two ORFs X (13 kDa) and Y (14 kDa) overlap ORF3 and encode proteins of unknown functions. CSSV, isolates of the species *Citrus yellow mosaic virus* (CYMV) and *Taro bacilliform virus* (TaBV) are the only badnaviruses which to date are known to code for more than 3 ORFs; CYMV and TaBV encodes respectively 6 and 4 ORFs [15, 28].

Apart from the symptom variability between mild and severe isolates [3, 4], little information is available about the variability of CSSV [25]. The first molecular variability study involved the analysis of the coat protein (CP) gene but was limited to different 1A isolates from Ghana [13]. A more extensive study of the molecular aspects of CSSV variability is now very relevant. Sequences of isolates originating from across the CSSV-endemic region including Togo, Ghana, Ivory Coast, and Nigeria are also required to improve and validate a PCR diagnostic test and allow better virus-indexing procedures. A better understanding of the genetic diversity of CSSV in West Africa and elsewhere will in turn help to provide a better understanding of the development of the epidemics and their eradication and of the evolution of viral populations. Obviously, the screening of new cocoa varieties for resistance against CSSV should furthermore take into account the virus variability present.

The first complete sequence of a CSSV isolate was determined in 1993 [9]. The tedious extraction process needed to obtain viral DNA from cacao leaves and the lack of alternate herbaceous hosts have been constraints in the study of other isolates. We present in this paper a PCR full length cloning strategy that has permitted the molecular cloning and complete sequencing of five other CSSV isolates, two originating from Togo and three originating from Ghana. The six full length sequences of CSSV have been aligned, the identity percentage between nucleotides and amino acids sequences of isolates has been calculated for the different ORFs and phylogenetic analyses have been made.

Methods

Virus isolates

CSSV isolates were collected as symptomatic leaves from infected cocoa tree (Nyongbo2 and Wobe12 respectively in the Agou and the Litime area in Togo, New Juaben, Peki and N1A in Ghana). Wobe12, Nyongbo2 and New Juaben are considered as severe isolates because they are associated with stem swelling symptoms. Peki and N1A are considered as mild isolates. Leaf samples were used fresh or frozen at -80°C .

Isolation of total plant DNA

DNA was extracted from fresh or frozen leaves, with the Dneasy[®] Plant Mini Kit (Qiagen) for each isolate. Fifty to 100 mg of material were frozen in liquid nitrogen and ground in a

microcentrifuge tube with a pestle. One hundred and fifty μ l of DNA solution was obtained after elution from the DNeasy column.

Primer design, PCR amplification

For Nyongbo2, New Juaben, Peki and N1A isolates, abutting primers Full-length+ (CGCTG-CAGTATTTC AAAGAA) and Full-length- (CTGCAGGAGCGTTTTTGA) were designed overlapping the *Pst*I restriction site (unique site for Agou1 isolate) in the RT domain of ORF3. For Wobe12 isolate, abutting primers FLWobeF (ATCGATGATATCCTAGTTTTTCAGC) and FLWobeR (GGATCGATGTAAACTGCAATGA) were designed overlapping the *Cla*I restriction site in the same domain as before. The primers were produced with 5' *Pst*I (or 5' *Cla*I) site to obtain cloned PCR fragments that could be excised with *Pst*I (or *Cla*I) to produce possibly infectious linear viral DNA.

PCR amplifications were performed in 50 μ l reaction mixture (reaction buffer supplied by the manufacturer Roche) containing 10 μ l of plant DNA extract, 2.25 mM MgCl₂, each dNTP at a concentration of 0.5 mM, each primer at a final concentration of 1.2 μ M and 2.6 U of Expand® High Fidelity thermostable DNA polymerase. Reactions were carried out in a Biometra T personal thermal cycler using an initial denaturation step of 92 °C for 2 min followed by 30 cycles of 92 °C for 10 s, 57 °C for 30 s and 68 °C for 8 min and a final elongation step of 68 °C for 7 min.

Cloning, sequencing

The resulting 7 kb PCR-amplified fragments of CSSV-Nyongbo2, -Wobe12, -NewJuaben, -Peki, and -N1A were cloned with a Topo® XL PCR cloning kit according to the manufacturer's instructions (Invitrogen). PCR amplified fragments were separated on 0.8% agarose gel stained with Crystal violet, excised from the gel and purified using a S.N.A.P. (Invitrogen) purification column supplied by the manufacturer. The fragment was inserted into the pCR-XL-Topo vector and then introduced into One Shot® TOP10 chemically competent cells. Recombinant plasmid DNA was purified using either the alkaline lysis method [26] or Qiaprep® spin miniprep kit (Qiagen). Clones were analysed by digestion with restriction endonucleases and one clone for each isolate was sent to Genome Express Corporation to be sequenced. Sequencing was done by primer walking. Three to six clones of each isolate were sequenced in a same region of 1.2 kb (including non-coding, ORF1 and ORF2 regions) to take into account the viral population diversity in a given plant. A CSSV-Nyongbo2 clone was amplified with Expand® High Fidelity thermostable DNA polymerase and the same primers as before. Three clones resulting for this sub-cloning were sequenced in a 1 kb region to estimate the level of sequence errors introduced during PCR.

Sequence analysis

The vector NTI Suite software was used to analyse the DNA sequence, as well to deduce and analyse amino acid sequence data. Phylogenetic relationships between CSSV isolates were estimated using the Darwin 4 program [22]. Amino acid sequences were aligned using the Clustal algorithm [27], and phylogenetic trees were constructed using a Neighbour-joining method on the basis of dissimilarities matrix. Robustness of the tree was determined by bootstrap sampling of the multiple alignment (1 000 sets).

Sequences of CSSV-Nyongbo2, CSSV-Wobe12, CSSV-New Juaben, CSSV-Peki and CSSV-N1A have been deposited in the EMBL Sequence Database as accession numbers AJ534983, AJ781003, AJ608931, AJ609019 and AJ609020 respectively. The Genbank accession numbers of sequences used for comparative purposes were AJ002234 (*Banana streak bacilliform virus* – BSV-Onne), X52938 (*Commelina yellow mottle virus* – CoYMV),

AF347695 (CYMV), L14546 (CSSV-Agou1), X94576 (*Dioscorea alata bacilliform virus* – DabV), M89923 (*Sugarcane bacilliform virus* – SCBV20), AJ277091 (SCBV-IM), AF357836 (*Taro bacilliform virus* – TaBV), AJ292232 (*Rice tungro bacilliform virus* – RTBV-AP), O76470 (RTBV-Serdang), X57924 (RTBV-Phil1).

Results

Cloning and restriction analysis of CSSV DNAs

Five to ten clones for each isolate were found to have an insert of around 7.0 kb on *Pst*I (or *Cla*I) digestion. For Peki, and New Juaben clones, the viral sequences were seen to have a second internal *Pst*I site.

Restriction maps, using restriction enzymes *Bam*HI, *Eco*RV and *Hind*III were constructed for one to three clones for each isolate and compared to that of Agou1. This enabled the determination of the sense of the insert, and differences were noticed between isolates and between clones of Peki isolate (data not shown). One clone for each isolate was chosen for sequencing.

Nucleotide sequence of the CSSV clones

Figure 1 presents the five new circular genomes compared that to Agou1. Numbering of the sequence follows badnavirus convention and refers to the plus-strand beginning at the 5' end of the minus-strand replication priming site. The length of the total genome ranges from 7006 bp (N1A) to 7297 bp (Wobe12).

The genomes of these newly sequenced isolates have G + C content of 43.5 to 44.3% which is very similar to that determined for the Agou1 isolate (44.1%).

Within the intergenic region, a possible TATA box (CTATAAATG between nucleotides 6961–6969 for CSSV Agou1) and a poly(A) signal (GAAATAAA between nucleotides 7066–7073 for CSSV-Agou1) were found highly conserved between the six CSSV isolates. They would allow the production of a terminally redundant full-length transcript.

In the intergenic region, the putative tRNA^{met} binding site sequence [11] can be found for all the new CSSV isolates in position 1–18 as in CSSV-Agou1. Fifteen to 16 nucleotides of 18 are complementary to the consensus sequence of plant tRNA^{met}. Apart from these conserved functional motifs, many insertion/deletions and substitution occurred in the intergenic region and nucleotide sequence identities of pairwise combinations of isolates range from 40.7% (between the Ghanaian isolate Peki and the Togolese isolate Agou1) to 93.1% (between the two Ghanaian isolates N1A and New Juaben). Percentage of nucleotide identity is surprisingly higher in the intergenic region (84.1%) than in the overall ORF3 coding region (74.4%) between Wobe12-CSSV and Agou1-CSSV. This obviously reflects the discontinuous and complex constraints which govern the evolution of these sequences.

Analysis of coding regions

The size, number and arrangement of Nyongbo2-CSSV ORFs are similar to Agou1 CSSV but the size, number and arrangement of ORFs in Wobe12-CSSV,

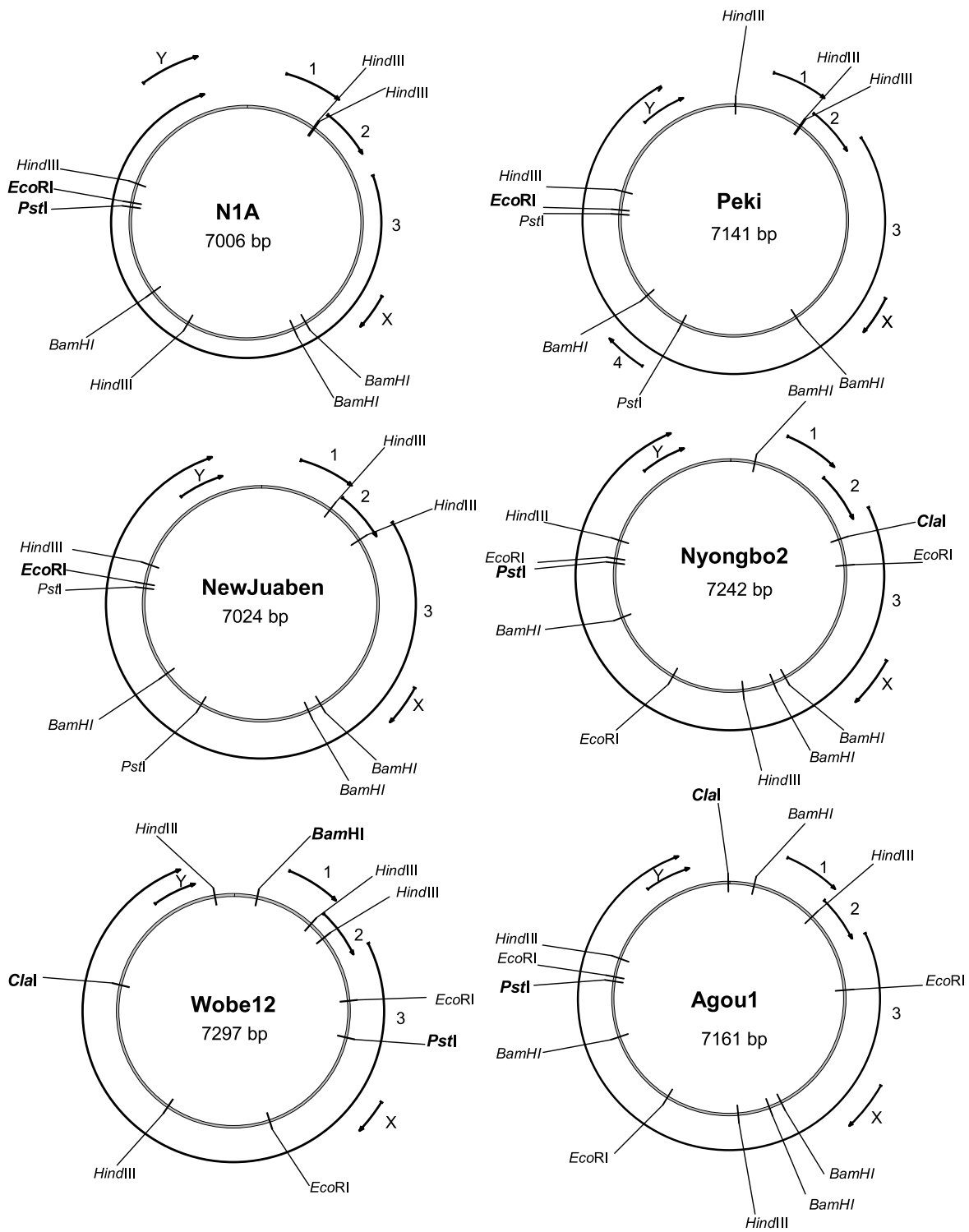


Fig. 1. Organisation of the circular genomes of new CSSV isolates compared to Agou1 (Hagen et al., 1993). Arrows indicate the deduced ORFs 1–4, X and Y capable of encoding proteins larger than 9 kDa. *BamHI*, *ClaI*, *EcoRI*, *HindIII* and *PstI* restriction sites are shown. Unique restriction sites are in bold

Table 1. Protein-coding regions located on the plus-strand of the genome of CSSV isolates

	Number of amino acids (<i>starting nucleotide-ending nucleotide*</i>) Calculated molecular mass (Da)						Sequence
	ORF1	ORF2	ORF3	ORFX	ORF4	ORFY	Size (bp)
Agou1	143 (441–869) 16755	132 (869–1264) 14443	1834 (1272–6773) 211385	113 (2374–2712) 13149	NC	131 (6434–6826) 14193	7161
Nyongbo2	143 (442–870) 16747	145 (870–1304) 15776	1839 (1273–6789) 211862	113 (2375–2713) 13191	NC	131 (6450–6842) 14250	7242
Wobe12	143 (432–860) 16754	149 (860–1306) 16381	1868 (1275–6878) 215278	81 (2458–2700) 9719	NC	131 (6563–6955) 14559	7297
New Juaben	143 (294–722) 16751	145 (722–1156) 15823	1847 (1125–6665) 212371	91 (2308–2580) 10867	NC	131 (6308–6700) 14167	7024
Peki	143 (296–724) 16752	145 (724–1158) 15852	1816 (1127–6574) 208995	90 (2310–2579) 10739	95 (4212–4496) 11058	131 (6307–6699) 14304	7141
N1A	143 (284–712) 16822	145 (712–1146) 15847	1770 (1345–6654) 203256	91 (2297–2569) 11007	NC	131 (6297–6689) 14276	7006

*Without including the stop codon

NC: No Corresponding ORF

N1A-CSSV, Peki-CSSV and New Juaben-CSSV are slightly different from Agou1-CSSV. Table 1 presents the protein-coding regions located on the plus-strand genome of CSSV isolates and capable of encoding proteins larger than 9 kDa. As was determined for CoYMV [19] we observed other putative ORFs on the minus strand genome of the three Ghanaian isolates but it is likely that these ORFs are not expressed on account of the generally accepted strategy of replication of the caulimoviridae [24].

ORF1 (429 nt) and ORFY (393 nt) sizes are conserved. ORF2 size (435 nt) is conserved among four isolates sequenced but not in Agou1 (396 nt) and in Wobe12 (447 nt). Differences in the size of ORF2 proteins between CSSV isolates correspond to a deletion of 13 amino acids at the C-terminus for Agou1 and to an insertion of 4 amino acids in the C-terminal domain for Wobe12. A nucleic acid-binding function was already demonstrated in the C-terminal domain of Agou1 ORF2 protein [16]. Analysis of the extension of the C-terminus of this protein in the other isolates showed the presence of other KPK motifs implicated in this nucleic acid-binding function. ORF3 varies from 5310 nucleotides (N1A) to 5604 nucleotides (Wobe12). Many insertion/deletions are detected in this ORF but majority of them concerned multiples of three nucleotides which do not affect the reading frame. The N-terminal 77 amino acids are deleted in

ORF3 polyprotein for N1A isolate. The amino acid sequence of ORF3 contains characteristic features of Caulimoviridae. One “cys” motif CXCX₂CX₄HX₄C and three CX₂C motifs are present for all newly sequenced isolates between amino acids 710 and 970 of ORF3 and are most probably involved in the RNA-binding function of the capsid protein. Consensus sequences for aspartyl protease (LDTGATT) and reverse transcriptase (YIDDIL) are conserved between the six isolates in C-terminal region of this polyprotein [9].

ORFX is 66 to 96 nucleotides smaller for the three Ghanaian isolates and Wobe12-CSSV compared to Agou1-CSSV and Nyongbo2-CSSV isolates, the putative proteins coded are translated in the same reading frame but from a different ATG initiator. Moreover, the protein coded by ORFX in the case of Wobe12 isolate is less than 10 kDa. The Peki-CSSV isolate contains an additional short ORF4 that encodes a putative protein of 11 kDa.

Arrangement of N1A-CSSV ORFs is not identical to other isolates; ORF1, ORF2/3 and ORFY are respectively coded in three different reading frames for this isolate compared to the five other isolates where ORF1/3 and ORF2/Y are respectively coded in two different open reading frames. For the six CSSV isolates, ORF2 overlaps ORF1 with a nucleotide frame change of -1 (ATGA: the termination codon of ORF1 is in italics and the initiation codon of ORF2 is underlined). This ATGA junction motif is also found in RTBV [12, 23] and in CYMV [15].

Table 2 shows the nucleotide and amino acid sequence percentage identity and similarity between the different ORFs of isolates. ORF1 is the most conserved ORF coding region (81 synonymous mutations versus 101 mutations among the 6 isolates) and the amino acid dissimilarity of pairwise combinations of CSSV isolates ranges from 1.4% to 8.4%. The maximum amino acid dissimilarity between isolates is 25.5% for ORF2, 23.8% for ORF3 and slightly higher for ORFY (33.6%). The value of overall variability calculated for ORF3 corresponds to a succession of more or less variable regions that can be identified on alignment. In agreement with the division of the ORF3 polyprotein in three regions [9, 12], we observed that the first region (amino acids 1–350) which correspond to the movement protein region is highly conserved as it was already observed for other pararetroviruses [21]. Region 2 is far less conserved (particularly amino acids 370–500 and 1010–1070). Region 3 has an intermediary level of variability. ORFX is the less conserved ORF. ORFX size and sequence are only conserved among the two Togolese isolates Agou1 and Nyongbo2 on one hand and among the three Ghanaian isolates on the other hand. The divergence observed between ORFX sequences of the different isolates was too high and they have not been aligned for a phylogenetic analysis.

Maximum nucleotide sequence variability between pairwise combinations of complete genomic sequences of CSSV isolates was 29.4% (between Wobe12-CSSV and Peki-CSSV).

Phylogenetic analysis

To determine the relationships between the six CSSV isolates, three phylogenetic trees were constructed using the deduced amino acid sequence of the ORF1, the

Table 2. Nucleotide sequence identity and amino acid sequence similarity of pairwise combinations of CSSV isolates for the ORFs 1, 2, 3, X and Y

	% of nucleotide sequence identity												% of amino acid sequence similarity																			
	ORF1				ORF2				ORF3				ORFX				ORFY															
	Ag1	Nb2	NJ	Peki	N1A	W12	Ag1	Nb2	NJ	Peki	N1A	W12	Ag1	Nb2	NJ	Peki	N1A	W12	Ag1	Nb2	NJ	Peki	N1A	W12								
Agou1	100	98.8	80	80.9	80.7	95.6	100	88.3	73.3	73.6	73.6	70.7	100	98.2	85.5	84.7	81.8	74.4	100	97.1	58.7	58.4	59.5	40.1	100	98.5	89.3	87.3	89.1	71		
	100	98.6	93.7	94.4	93	97.2	100	89	82.8	82.1	82.1	74.5	100	98.2	90.6	90.2	86.5	79.8	100	92.9	41.9	45.3	42.7	22.4	100	97.7	92.4	87.8	89.3	69.5		
Nyongbo2	100	80.2	81.1	80.9	94.9	100	100	81.8	81.8	77.6	100	85.1	84.4	81.5	74.2	100	58.7	58.1	60.1	41.3	100	88.3	86.3	88	71	100	90.1	85.5	87	71		
	100	93.7	94.4	93	97.2	100	100	89.7	88.3	88.3	76.5	100	90.3	89.8	86.3	79.7	100	41.9	43.6	44.4	22.4	100	90.1	85.5	87	71	100	90.1	85.5	87	71	
New	100	97.2	96.7	80.9	100	96.6	96.3	76.7	100	96.4	93.3	73.3	100	96.2	93.6	79.5	96	96.3	47.3	100	92.9	94.9	71.8	100	90.1	92.4	69.5	100	90.1	92.4	69.5	
Juaben	100	97.9	97.9	92.3	100	96.6	96.6	77.9	100	96.6	96.6	76.5	100	96.2	93.6	79.5	100	96.4	93.3	73.3	100	92.9	94.9	71.8	100	90.1	92.4	69.5	100	90.1	92.4	69.5
Peki	100	97.7	81.4	100	96.6	76.5	100	96.6	76.5	100	91.8	73.5	100	94.1	46.9	100	100	94.1	46.9	100	92.9	94.9	71.8	100	90.1	92.4	69.5	100	90.1	92.4	69.5	
	100	97.2	93.7	100	97.2	93.7	100	95.9	77.2	100	92	80	100	92	80	100	100	92	80	100	92.9	94.9	71.8	100	90.1	92.4	69.5	100	90.1	92.4	69.5	
N1A	100	80.7	100	100	100	77	100	77	100	70.4	100	51.3	100	100	51.3	100	100	51.3	100	20.5	100	92.9	94.9	71.8	100	90.1	92.4	69.5	100	90.1	92.4	69.5
	100	91.6	100	100	100	78.5	100	78.5	100	76.2	100	20.5	100	20.5	100	20.5	100	20.5	100	20.5	100	92.9	94.9	71.8	100	90.1	92.4	69.5	100	90.1	92.4	69.5
Wobe12	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	
	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	

Ag1 = Agou1; Nb2 = Nyongbo2; NJ = New Juaben; W12 = Wobe12

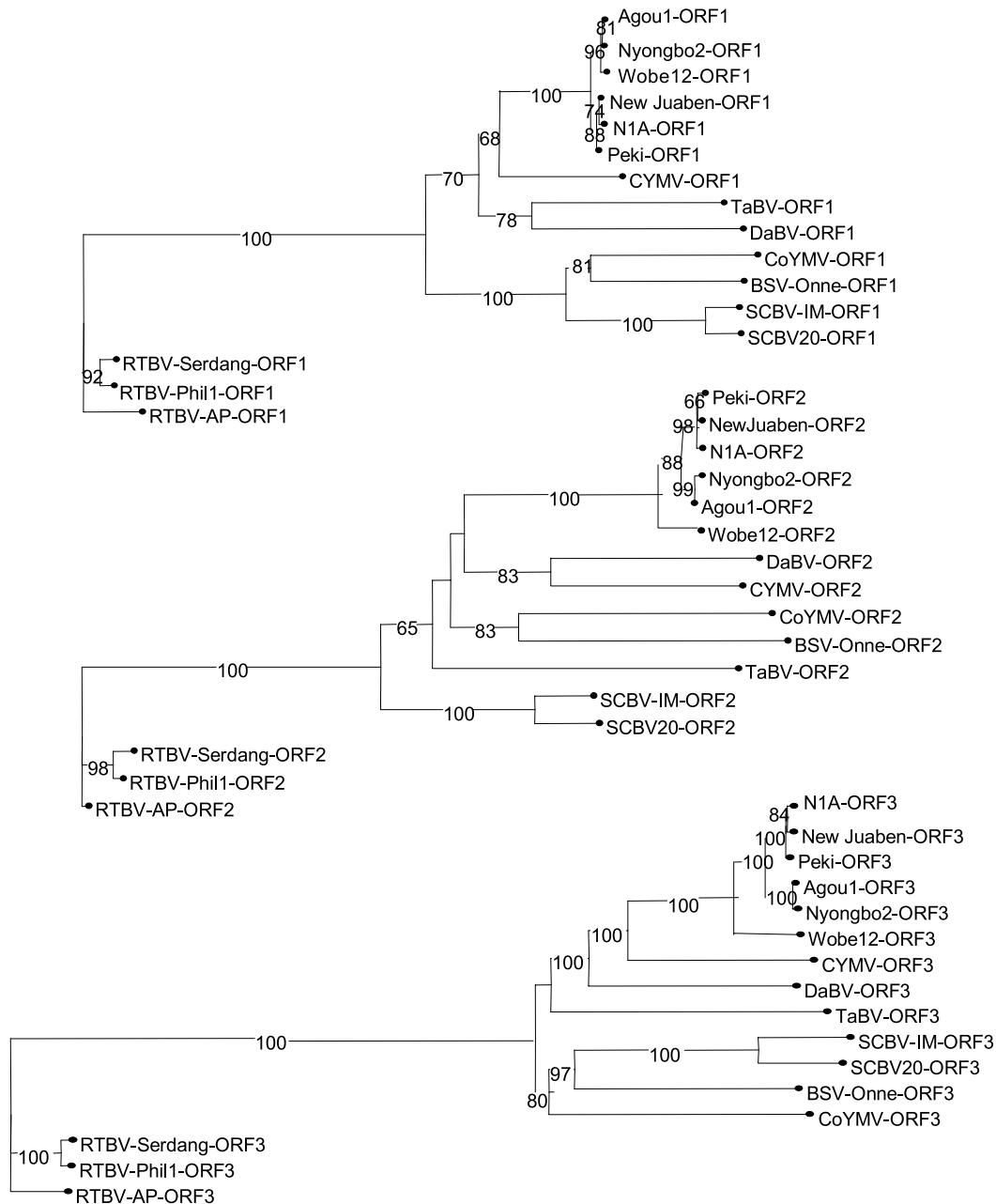


Fig. 2. Phylogenetic neighbour-joining trees generated by the Darwin 4 program based on predicted amino acids sequences of ORFs 1, 2 and 3. Numbers at the nodes of the branches represent percentage bootstrap values (1000 replicates) when superior to 60

ORF2, and the ORF3 proteins (Fig. 2) and one phylogenetic tree was deduced from complete nucleotide sequences alignment (Fig. 3). Sequences of corresponding ORFs of BSV-Onne, CoYMV, CYMV, DaBV, SCBV20, SCBV-IM, TaBV, RTBV-AP, RTBV-Phil1 and RTBV-Serdang were included in the phylogenetic analyses. As ORFY is present only in CSSV (and in a similar position for

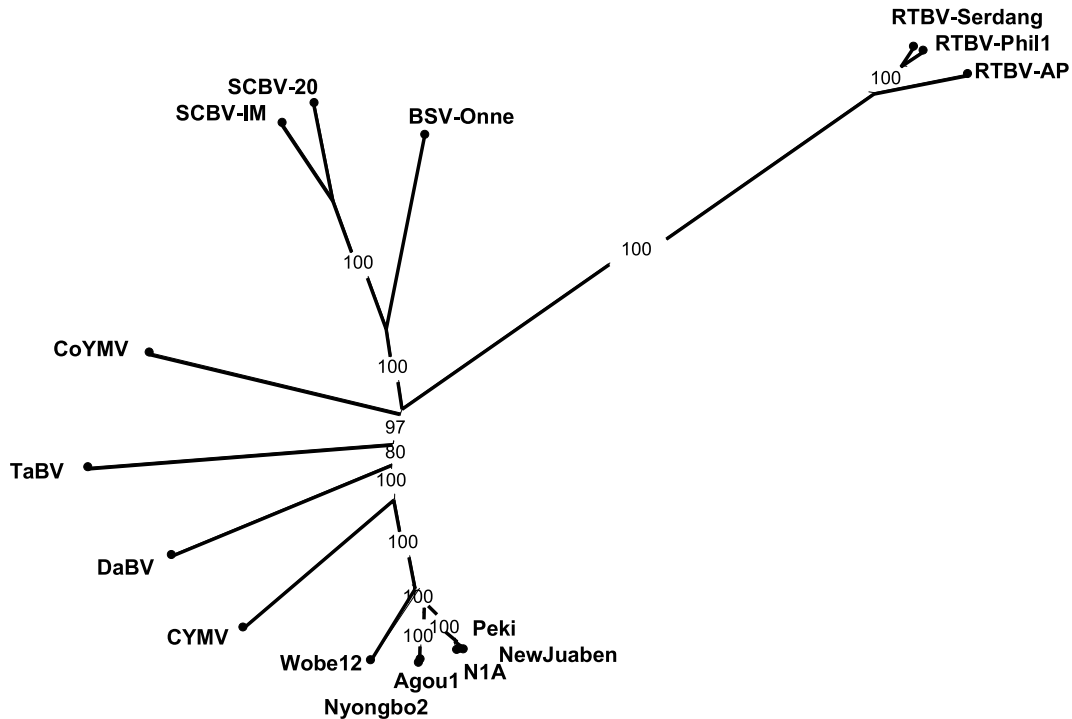


Fig. 3. Neighbour-joining tree generated by the Darwin 4 program based on complete nucleotide sequences of badnaviruses and RTBV. Numbers at the nodes of the branches represent percentage bootstrap values (1000 replicates) when superior to 60

CYMV-ORF6), a phylogenetic tree has not been done for this ORF. Irrespective of the ORF analysed, the results showed that Togolese isolates branched separately from Ghanaian isolates. Except for the phylogenetic tree constructed using the amino acid sequence of the ORF1 proteins, the different phylogenetic trees made a clear distinction between Wobe12-CSSV isolate from Litime area and the two other CSSV isolates from Agou area. This Togolese isolate from Litime is equally distant to the two other CSSV groups as confirmed by the amino acid and nucleotide identities (Table 2).

In all these analysis, CSSV isolates are closer to each other than to other badnaviruses and to the RTBV-like genus. The maximum variability level observed between the six CSSV isolates is slightly higher than the one observed between the three RTBV isolates [21].

Bootstrap values indicate the robustness of the grouping of Agou1-CSSV and Nyongbo2-CSSV isolates together, separate from a group composed of New-Juaben-, Peki- and N1A-CSSV isolates together and from another group composed of Wobe12-isolate. The highest value bootstrap scores are obtained for ORF3 amino acids sequences and complete nucleotide sequences.

Discussion

We demonstrate here the usefulness of the PCR full-length cloning strategy replacing the standard and time-consuming protocols to obtain new CSSV clones

from small quantities of total nucleic acids extracted from diseased plants. PCR as a method for producing full-length, infectious clones from small tissue samples was reported for geminiviruses [1]. The abutting primers included a *Pst*I or *Cla*I site in order to allow the recovery of cloned amplification products from appropriately digested vectors. Infectivity of such full-length clones could be tested using microprojectile bombardment of cocoa seeds [10].

To date, CSSV-Agou1(7161 pb) has the smallest genome reported amongst the genus *Badnavirus*, but that of N1A-CSSV (7006 pb) is even smaller.

A comparison of the complete ORF3 between the six CSSV isolates revealed that nucleotide sequence identities ranged from 70.4 to 98.2% (Table 2). The ORF3 variability was shown to be representative of the overall variability between the six isolates, which could be expected as ORF3 accounts for ~75% of the total genome. The overall ORF3 variability is slightly lower than the variability found in the ORF3 RT domain between different BSV isolates [7] as well as between CSSV isolates from different parts of Togo (60% identity, data not published) but this domain is probably neither the most conserved domain of a given badnavirus nor the most representative of its complete sequence variability. There is still a need to study complete sequences of other CSSV isolates to understand the genomic diversity of CSSV but these results are very useful to improve our understanding of the variable and conserved regions along the genome of CSSV. This has allowed the determination of new primers in conserved regions like the first region of ORF3 for a more versatile diagnosis than with primers Badna 2.1/3.1CSSV located in the RT region of ORF3 [20]. This type of information could equally permit to choose different regions instead of the RT domain for a more extensive study of molecular diversity of badnaviruses to improve their taxonomy.

Relationships between isolates are quite similar for nucleotide or amino acid sequences of the four ORFs, with the exception of ORF1 showing a different pattern where all Togolese isolates branched together. This gene indeed seems to be under a different selection pressure (percent of amino acid similarity slightly higher than for ORFs 2, 3 and Y). CSSV isolates can be separated into three groups, one of which containing Wobe12 isolate more distantly related to the others. Phylogenetic relationships between Ghanaian and Togolese isolates are more influenced by their geographical origin than on whether the sequences originate from mild or severe isolates. More isolates from different regions should now be analysed to confirm this result.

Sequence variability was analysed between clones of the same isolate in non-coding and ORF1-2 regions. This variability ranging from 0 to 0.7%, was found to be similar to the variability found between two CoYMV clones [19] but lower than the 4.2% of variability found between two DaBV clones [2]. In order to prove that this variability is not due to errors introduced by the Expand® High Fidelity thermostable DNA polymerase used for amplification, another amplification experiment was carried out using a Nyongbo2 full-length clone preparation with the same enzyme source. Cloning was done with the same vector and four clones from this experiment were sequenced. No variation was observed between these clones compared to those obtained from PCR full-length amplification of a crude plant DNA extract. The sequence variability observed should therefore be

considered to reflect true intra-plant virus variability. It should be noted that the low variability found between clones indicates that CSSV populations could be more homogeneous compared to other badnaviruses as BSV and SCBV for which a high heterogeneity has been reported [7, 8].

Surprisingly, we observed one base change both for Nyongbo2 and Peki isolates at two different positions in the 12 nucleotide block TGGTATCAGAGC corresponding to the putative tRNA^{met} binding site that is conserved among all the pararetrovirus isolates sequenced to date [15, 14]. This change is observed for the full sequenced clones of these isolates but not for the five other partially sequenced clones of these isolates. Because a mutation in this binding site would most probably be lethal due to hindering viral replication, the presence of this variability may support the concept of the viral quasi-species with coexistence in a population of molecules with different characteristic and different fitness [5, 6].

The large difference between ORFX of Nyongbo2-CSSV and Agou1-CSSV isolates and the four other isolates could be an indication that the putative protein coded is fortuitous and not relevant for viral replication and pathogenicity. This is supported by the fact that ORFX found in the Ghanaian isolates encodes for a putative protein 22 to 23 amino acids smaller and very different from the one coded by the two Togolese Nyongbo2 and Agou1 isolate. Indeed, three discontinuities between these sequences can be observed, with the initiation taking place at a different position and the amino acid similarity ranging from only 41.9 to 45.3%. Moreover, the short ORFX present on Wobe12-CSSV sequence code for a putative protein far distantly related with the putative proteins coded by ORFX of the other isolates (from only 14.3% to 22.4% of amino acid similarity which has not any biological significance). The ORFX region is clearly under relaxed selection pressure compared to the other ORFs under negative selection pressure [29, 6]. No proteins were found to be similar to the putative product of Peki-CSSV ORF4 by BlastP search, and this short ORF is only present in one isolate. These facts could equally suggest a subordinate role for the putative protein encoded. Further studies including construction of mutants and a test of their effects are needed to test this hypothesis. In contrast, ORFY is surprisingly conserved amongst the five CSSV-isolates despite the unknown function of the protein encoded.

References

1. Briddon RW, Prescott AG, Lunness P, Chamberlin LCL, Markham PG (1993) Rapid production of full-length, infectious geminivirus clones by abutting primer PCR (AbP-PCR). *J Virol Methods* 43: 7–20
2. Briddon RW, Phillips S, Brunt A, Hull R (1999) Analysis of the sequence of *Dioscorea alata bacilliform virus*: comparison to others members of the badnavirus group. *Virus Genes* 18: 277–283
3. Castel C, Amefia YK, Djiekpor EK, Partiot M, Segbor A (1980) Le swollen shoot du cacaoyer au Togo. Les différentes formes de viroses et leurs conséquences économiques. *Café Cacao Thé* 24: 131–146
4. Dufour B (1988) Utilisation d'une méthode de transmission pour caractériser les formes togolaises du virus du cacao swollen shoot du cacaoyer. *Café Cacao Thé* 32: 219–228

5. Garcia-Arenal F, Palukaitis P, Zaitlin M (1984) Strains and mutants of *Tobacco mosaic virus* are both found in virus populations. *Virology* 132: 131–137
6. Garcia-Arenal F, Fraile A, Malpica JM (2001) Variability and genetic structure of plant virus populations. *Annu Rev Phytopathol* 39: 157–186
7. Geering ADW, McMichael LA, Dietzgen RG, Thomas JE (2000) Genetic diversity among *Banana streak virus* isolates from Australia. *Phytopathology* 90: 921–927
8. Geijskes RJ, Braithwaite KS, Dale JL, Harding RM, Smith GR (2002) Sequence analysis of an Australian isolate of sugarcane bacilliform badnavirus. *Arch Virol* 147: 2393–2404
9. Hagen LS, Jacquemond M, Lepingle A, Lot H, Tepfer M (1993) Nucleotide sequence and genomic organization of cacao swollen shoot virus. *Virology* 196: 619–628
10. Hagen LS, Lot H, Godon C, Tepfer M, Jacquemond M (1994) Infection of *Theobroma cacao* using cloned DNA of cacao swollen shoot virus and particle bombardment. *Mol Plant Pathol* 84: 1239–1243
11. Harper G, Hull R (1998) Cloning and sequence analysis of banana streak virus. *Virus Genes* 17: 271–278
12. Hay JM, Jones MC, Blakebrough ML, Dasgupta I, Davies JW, Hull R (1991) An analysis of the sequence of an infectious clone of rice tungro bacilliform virus. *Nucleic Acids Res* 19: 2615–2621
13. Hoffmann K, Sackey ST, Maiss E, Adomako D, Vetten HJ (1997) Immunocapture polymerase chain reaction for the detection and characterization of cacao swollen shoot virus 1A isolates. *J Phytopathol* 145: 205–212
14. Hohn T, Hohn B, Pfeiffer P (1985) Reverse transcription in CamV. *Trends Biochem Sci* 10: 205–209
15. Huang Q, Hartung JS (2001) Cloning and sequences analysis of an infectious clone of *Citrus yellow mosaic virus* that can infect sweet orange via *Agrobacterium*-mediated inoculation. *J Gen Virol* 82: 2549–2558
16. Jacquot E, Hagen LS, Jacquemond M, Yot P (1996) The open reading frame 2 product of cacao swollen shoot badnavirus is a nucleic acid-binding protein. *Virology* 225: 191–195
17. Lockhart BEL (1990) Evidence for a double-stranded circular DNA genome in a second group of plant viruses. *Phytopathology* 80: 127–131
18. Lot H, Djiekpor E, Jacquemond M (1991) Characterization of the genome of cacao swollen shoot virus. *J Gen Virol* 72: 1735–1739
19. Medberry SL, Lockhart BEL, Olszewski NE (1990) Properties of *Commelina* yellow mottle virus's complete DNA sequence, genomic discontinuities and transcript suggest that it is a pararetrovirus. *Nucleic Acids Res* 18: 5505–5513
20. Muller E, Jacquot E, Yot P (2001) Early detection of cacao swollen shoot virus using the polymerase chain reaction. *J Virol Methods* 93: 15–22
21. Nath N, Mathur S, Dasgupta I (2002) Molecular analysis of two complete rice tungro bacilliform virus genomic sequences from India. *Arch Virol* 147: 1173–1187
22. Perrier X, Flori A, Bonnot F (2003) Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Genetic diversity of cultivated tropical plants. Enfield Science Publishers, Montpellier, pp 43–76
23. Qu RD, Bhattacharyya M, Laco GS, De Kochko A, Rao BL, Kaniewska MB, Elmer JS, Rochester DE, Smith CE, Beachy RN (1991) Characterization of the genome of rice tungro bacilliform virus: comparison with *Commelina* yellow mottle virus and caulimoviruses. *Virology* 185: 354–364 [published erratum *Virology* 1992 Feb;186(2):798]
24. Rothnie HM, Chapdelaine Y, Hohn T (1994) Pararetroviruses and retroviruses: a comparative review of viral structure and gene expression strategies. *Adv Virus Res* 44: 1–67

25. Sagemann W, Lesemann DE, Paul HL, Adomako D, Owusu GK (1985) Detection and comparison of some ghanaiian isolates of cacao swollen shoot virus (CSSV) by enzyme-linked immunosorbent assay (ELISA) and immunoelectron microscopy using an antiserum to CSSV strain 1A. *Phytopathol Z* 114: 79–89
26. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning : A laboratory manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
27. 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
28. Yang IC, Hafner GJ, Dale JL, Harding RM (2003) Genomic characterisation of taro bacilliform virus. *Arch Virol* 148: 937–949
29. Zhang J (2003) Evolution by gene duplication : an update. *Trends Ecol Evol* 18: 292–298

Author's address: Emmanuelle Muller, CIRAD, UMR BGPI, TA 41/K, Campus International de Baillarguet, 34398 Montpellier cedex 5, France; e-mail: emmanuelle.muller@cirad.fr