Queensland University of Technology
Brisbane Australia

This may be the author's version of a work that was submitted/accepted for publication in the following source:

James, Anthony, Geijskes, Robert, Dale, James, & Harding, Rob
(2011)
Molecular characterisation of six badnavirus species associated with leaf streak disease of banana in East Africa.
Annals of Applied Biology, *158*(3), pp. 346-353.

This file was downloaded from: https://eprints.qut.edu.au/39039/

© Consult author(s) regarding copyright matters

This work is covered by copyright. Unless the document is being made available under a Creative Commons Licence, you must assume that re-use is limited to personal use and that permission from the copyright owner must be obtained for all other uses. If the document is available under a Creative Commons License (or other specified license) then refer to the Licence for details of permitted re-use. It is a condition of access that users recognise and abide by the legal requirements associated with these rights. If you believe that this work infringes copyright please provide details by email to qut.copyright@qut.edu.au

Notice: *Please note that this document may not be the Version of Record (i.e. published version) of the work. Author manuscript versions (as Submitted for peer review or as Accepted for publication after peer review) can be identified by an absence of publisher branding and/or typeset appearance. If there is any doubt, please refer to the published source.*

https://doi.org/10.1111/j.1744-7348.2011.00466.x

MOLECULAR CHARACTERISATION OF SIX BADNAVIRUS SPECIES ASSOCIATED WITH LEAF STREAK DISEASE OF BANANA IN EAST AFRICA

A. P. James, R. J. Geijskes, J. L. Dale and R. M. Harding

Centre for Tropical Crops and Biocommodities, Queensland University of Technology, GPO Box 2434, Brisbane, Qld, 4001, Australia.

19

20 **ABSTRACT**

21 Banana leaf streak disease, caused by several species of Banana streak virus (BSV), is
22 widespread in East Africa. We surveyed for this disease in Uganda and Kenya, and used
23 rolling-circle amplification (RCA) to detect the presence of BSV in banana. Six distinct
24 badnavirus sequences, three from Uganda and three from Kenya, were amplified for which
25 only partial sequences were previously available. The complete genomes were sequenced
26 and characterised. The size and organisation of all six sequences was characteristic of other
27 badnaviruses, including conserved functional domains present in the putative polyprotein
28 encoded by open reading frame (ORF) 3. Based on nucleotide sequence analysis within the
29 reverse transcriptase/ribonuclease H-coding region of open reading frame 3, we propose
30 that these sequences be recognised as six new species and be designated as *Banana streak*
31 UA *virus*, *Banana streak* UI *virus*, *Banana streak* UL *virus*, *Banana streak* UM *virus*, *Banana*
32 *streak* CA *virus* and *Banana streak* IM *virus*. Using PCR and species-specific primers to test
33 for the presence of integrated sequences, we demonstrated that sequences with high
34 similarity to BSIMV only were present in several banana cultivars which had tested negative
35 for episomal BSV sequences.

36

37 **INTRODUCTION**

38 Bananas (*Musa* spp) are hosts to several badnaviruses collectively named banana streak
39 virus (BSV, genus *Badnavirus*, family *Caulimoviridae*). BSV infection causes leaf streak
40 disease which is characterised by distinct chlorotic and necrotic flecking on leaves, as well as
41 a range of other symptoms including distortion of leaves and petioles, stem cracking,
42 abnormal bunch development and death of the growing point (Dahal *et al*., 2000). Although
43 found in most banana growing regions, leaf streak is the most frequently observed viral
44 disease of banana in the Americas and most of Africa (Geering, 2009; and references
45 therein). The disease is particularly widespread in both Uganda (Harper *et al*., 2004, 2005)
46 and Kenya (Karanja *et al*., 2008).

47 Badnaviruses are plant pararetroviruses with non-enveloped, bacilliform-shaped
48 virions of approximately 30 x 130-150 nm, and circular double-stranded DNA genomes of 7-
49 8 kbp (Hull *et al*., 2005). All badnaviruses typically encode three open reading frames (ORFs)
50 on the virus-sense coding strand. ORF 1 encodes a small protein of unknown function which
51 has been associated with virions (Cheng *et al*., 1996). ORF 2 encodes a protein of ~14 kDa
52 with a non-specific DNA- and RNA-binding activity (Jacquot *et al*., 1996). This protein may
53 function in virion assembly due to the presence of a predicted N-terminal coiled-coil domain
54 which supports self-interaction to form a tetramer (Leclerc, 1998). The large polyprotein
55 encoded by ORF 3 contains domains associated with movement, the virus capsid, aspartic
56 protease, reverse transcriptase (RT) and ribonuclease H (RNaseH) functions.

57 Three distinct species of BSV, namely *Banana streak* OL *virus* (BSOLV), *Banana streak*
58 GF *virus* (BSGFV) and *Banana streak* MY *virus* (BSMYV), previously *Banana streak Mysore*
59 *virus*, are now recognised by the International Committee on the Taxonomy of Viruses

60 (ICTV) (Geering, 2010), while a fourth species, *Banana streak* VN *virus* (BSVNV, previously
61 *Banana streak acuminata Vietnam virus*) has also recently been proposed based on full-
62 length sequence analyses (Lheureux et al., 2007). However, many other BSV species are
63 thought to exist based on the analyses of numerous reported partial sequences. Geering *et*
64 *al*., (2000; 2005A) reported partial sequences of two BSV isolates from Australia (named
65 Banana streak Cavendish virus (BSV-Cav) and Banana streak Imové virus (BSImV)), while
66 Harper *et al*., (2005) reported the presence of 13 distinct BSV sequence groups from
67 Uganda, named consecutively as Banana streak Uganda A virus to Banana streak Uganda M
68 virus. Further, a complete but as yet unpublished BSV sequence, named Banana streak
69 acuminata Yunnan virus, has been deposited in the NCBI database (GenBank accession no.
70 DQ092436). This isolate is phylogenetically most closely related to BSVNV (Gayral and Iskra-
71 Caruana, 2009).

72 Several episomal BSV sequences, namely BSOLV, BSMYV, BSGFV and BSIMV, have
73 been shown to have integrated counterparts, termed endogenous BSV (eBSV) , in the *Musa*
74 genome (Harper *et al*., 1999; Geering *et al*., 2005A & B; Gayral *et al*., 2008). Under certain
75 stress conditions, these sequences can be activated to cause episomal infections (Ndowora
76 *et al*., 1999; Dallot *et al*., 2001; Côte *et al*., 2010). Although many other endogenous
77 badnavirus sequences occur in the banana genome, these have no known episomal
78 counterparts and are not known to give rise to episomal infections (Geering *et al*., 2005A).
79 While endogenous badnaviruses occur in genetic backgrounds which include both *Musa*
80 *acuminata* (A-genome) and *M*. *balbisiana* (B-genome) and their hybrids, eBSVs are only
81 known to occur in some *Musa* accessions which contain a B-genome.

82 As part of a Grand Challenges in Global Health initiative funded by the Bill and
83 Melinda Gates Foundation, we have been developing a diagnostic capacity for banana
84 viruses in East Africa with a specific focus on BSV. Diagnostic tests for BSV have been
85 complicated by the extensive genetic and serological diversity that exists amongst BSV
86 isolates and the presence of integrated BSV sequences in some banana cultivars which leads
87 to false positives. Recently, however, we have developed a rolling-circle amplification (RCA)-
88 based assay that specifically detects episomal, and not integrated, BSV sequences (James *et*
89 *al*., 2011). To determine the prevalence of BSV in East Africa, we conducted disease surveys
90 of bananas in Uganda and Kenya and tested samples by RCA. In this paper, we report for the
91 first time the complete nucleotide sequence and molecular characterisation of six distinct
92 BSV species from these two countries.

93

94 **MATERIALS AND METHODS**

95 Virus nomenclature

96 For consistency throughout the remainder of this manuscript, we have elected to use the
97 nomenclature suggested by Geering (2010) to describe BSV species. Banana streak Imové
98 virus (BSImV; Geering *et al*., 2005A) will be referred to as *Banana streak* IM *virus* (BSIMV);
99 Banana streak Cavendish virus (BSV-Cav; Geering *et al*., 2000) will be referred to as *Banana*
100 *streak* CA *virus* (BSCAV); Banana streak Uganda A virus (BSUgAV; Harper *et al*., 2005) will be
101 referred to as *Banana streak* UA *virus* (BSUAV); Banana streak Uganda I virus (BSUgIV;
102 Harper *et al*., 2005) will be referred to as *Banana streak* UI *virus* (BSUIV); Banana streak
103 Uganda L virus (BSUgLV; Harper *et al*., 2005) will be referred to as *Banana streak* UL *virus*

104 (BSULV); and Banana streak Uganda M virus (BSUgMV; Harper *et al*., 2005) will be referred

105 to as *Banana streak* UM *virus* (BSUMV).

106

107 Plant samples and amplification of viral DNA

108 Leaf samples were collected from banana plants displaying the chlorotic and necrotic

109 flecking symptoms typically associated with leaf streak disease (Table 1). Three samples

110 (Ug1, Ug8 & Ug12) were collected in south-west Uganda during a survey conducted in April

111 2008, and were previously shown to contain BSV-like sequences with homology to BSUIV,

112 BSULV and BSUMV, respectively (James *et al*., 2011). One sample (Ke171) was collected in

113 western Kenya during a survey in April 2009, while the remaining two samples (Ke8 & Ke10)

114 were obtained from the Kenyan Agricultural Research Institute research station at Njoro,

115 Kenya. Total nucleic acid (TNA) extracts were prepared and virus DNA amplified using the

116 Illustra TempliPhi 100 Amplification Kit (GE Healthcare, Buckinghamshire, United Kingdom)

117 as described previously (James *et al*., 2011).

118 Leaf samples were also obtained from 12 genotypically diverse *Musa* cultivars (Table

119 2) growing in tissue culture at DEEDI, Agri-Science Queensland, Nambour, Australia. These

120 plants had previously been certified as BSV negative using immuno-sorbent electron

121 microscopy (ISEM, Geering *et al*., 2000) as well as by RCA (James *et al*., 2011).

122

123

124

125

126

127

128

129

130

Table 1 Plant samples for RCA analysis and PCR primers used for detection of integrated sequences

Sample	Primer 1 (5´ - 3´)	Primer 2 (5´ - 3´)	Amplicon size (bp)	Annealing temp (°C)	Region of genome amplified
Ug1	GAACTGACAGTAGCGCAATCG	GACTTGGCTTGCCTGAGTATCG	943	60	6282-7224
Ug8	GAATCCTCAAAGGTACCCC	CATGAGGTCAAGCATATGC	619	50	435-1053
Ug12	GACGAGCTGCAAGCTCTCAGG	TGTGCCTATTCTGAGGTTGG	467	50	973-1439
Ke8	CTCAGCGGCAAGATTAGGAAGG	TCCCCATTGGTCGTCATTGC	517	60	6513-7029
Ke10	GCTAGGAAGAAAAGTCTGGG	TGCAAGTCTACTTACACAGC	475	50	7417-122
Ke171	AGGATTGGATGTGAAGTTTGAGC	ACCAATAATGCAAGGGACGC	783	57	6425-7207

131

132

133

134

135

136

137

Table 2 BSV-indexed leaf samples used for PCR analysis

Cultivar	Genotype	Virus specific PCR test					
		BSUAV	BSUIV	BSULV	BSUMV	BSCAV	BSIMV
Calcutta 4	AA	-	-	-	-	-	-
Pisang Oli	AA	-	-	-	-	-	-
Yangambi km5	AAA	-	-	-	-	-	-
NC-301	AAA	-	-	-	-	-	-
FHIA-17	AAAA	-	-	-	-	-	-
Da Jiao	ABB	-	-	-	-	-	-
Ainu	AAB	-	-	-	-	-	+
SH-3460.10	AAAB	-	-	-	-	-	-
FHIA-03	AABB	-	-	-	-	-	+
Balonkawe	ABB	-	-	-	-	-	+
Goly Goly Pot Pot	ABB	-	-	-	-	-	+
Lal Velchi	BB	-	-	-	-	-	-

138

139

140 Cloning and sequencing of virus DNA

141 RCA-amplified virus DNA was digested using either *Stu*I or *Pst*I and products were separated
142 by agarose gel electrophoresis. DNA was subsequently cloned into pUC19 and sequenced as
143 described previously (James *et al*., 2011). In all cases, at least three independent clones
144 were sequenced in both directions to determine a consensus sequence. Putative
145 identification of cloned fragments was made by comparison to sequences in NCBI database
146 (http://www.ncbi.nlm.nih.gov) using the Basic local alignment search tool (BLAST)
147 programs. According to the ICTV criteria (Hull *et al*., 2005) for species demarcation within
148 the genus *Badnavirus*, sequence differences within the RT/RNaseH-coding region of more
149 than 20% are considered distinct badnavirus species. Sequence comparisons were based on
150 a 529 bp region of the RT/RNaseH-coding region delimited by the BadnaFP/RP primers
151 reported in Yang *et al*. (2003). In cases where BSV isolates were identified for which full-
152 length sequences had not been previously reported, the complete sequence was obtained
153 by primer-walking. The sequence spanning putative restriction sites was confirmed by
154 sequencing of PCR products generated using sequence-specific primers for each putative
155 site present in each virus isolate. PCR mixes (20 µL) contained 10 µl 2x GoTaq Green Master
156 Mix (Promega Corp, Madison, WI), 5 ρmol of each primer, 1 µl of nucleic acid extract and
157 water to final volume. PCR cycling conditions were an initial denaturation of 94°C for 2 min
158 followed by 35 cycles of 94°C for 20 s, 50°C for 20 s, and 72°C for 30 s, with a final extension
159 at 72°C for 2 min. Reactions products were electrophoresed through 1.5% agarose gels,
160 stained using 0.25x SYBR® Safe DNA Gel Stain (Invitrogen Corp, Carlsbad, CA) and DNA
161 fragments visualised on a Safe imager blue-light transilluminator. Amplified fragments were

162

163 cloned into pGEM-T easy (Promega Corp, Madison, WI) and three clones were sequenced in
164 both directions using universal M13 primers.

165

166 PCR for detection of integrated sequences

167 PCR primers for detection of each virus were designed from sequences obtained in this
168 study (Table 1), with the exception of BSCAV, for which previously published primers were
169 used (Geering *et al*., 2000). PCR was carried out as described above, with annealing
170 temperatures of 50°C, 57°C or 60°C (Table 1).

171

172 Phylogenetic analysis

173 Full-length sequences were assembled in VectorNTI Advance v.11 (Invitrogen Corp,
174 Carlsbad, CA), which was subsequently used for the identification of putative ORFs and
175 other conserved genomic features of badnaviruses, as well as translation of putative ORFs
176 for analysis of conserved domains present in badnavirus proteins. Conserved features of
177 genome sequences were identified as previously described (Geering *et al*., 2010; Lheureux
178 *et al*., 2007). For phylogenetic analysis and sequence comparison to published badnavirus
179 sequences, RT/RNaseH core sequences were identified using the method of Geering *et al*.,
180 (2010). RT/RNaseH sequences were aligned using the CLUSTAL W algorithm in MEGA 4.0
181 (Tamura *et al*., 2007). Phylogenetic trees were constructed using the neighbour-joining

182 method, following pair-wise sequence comparison using the Kimura 2-parameter model in
183 MEGA 4.0 (Kimura, 1980).

184

185 **RESULTS**

186 Cloning and sequencing

187 Nucleic acid was extracted from the three Ugandan leaf samples (Ug 1, 8 and 12) and three
188 Kenyan leaf samples (Ke8, 10 and 171) and was subjected to RCA. The products were
189 digested with *Stu*I (or *Pst*I in the case of sample Ke171) and analysed by agarose gel
190 electrophoresis. A single band of ~7.5 kbp, presumably representing full-length BSV genomic
191 DNA, was observed in the Ug1, Ke8 and Ke171 extracts, while two bands (~5.5 and 1.8 kbp),
192 three bands (~3.7, 2.2 and 1.6 kbp) and four bands (~7.5, 4.4, 2.1 and 1 kbp) were observed
193 in the Ug8, Ug12 and Ke10 extracts, respectively.

194 When the restriction fragment/s from Ug1, Ug12, Ke8 and Ke171 were cloned and
195 analysed, contiguous full-length sequences were obtained. Analysis of the RT/RNaseH-
196 coding region of these sequences revealed 82-86%, 90-95%, 94% and 94-95% similarity to
197 BSULV, BSUMV, BSUAV and BSCAV isolates, respectively. Analysis of the 3' and 5' terminal
198 sequences of the ~7.5 kbp band derived from Ke10 revealed it was identical to that of Ke8.
199 When the remaining three restriction fragments from this isolate were cloned and analysed,
200 a contiguous full-length sequence was obtained. The sequence of the RT/RNaseH-coding
201 region showed 97-99% similarity to BSIMV.

202 Although analysis of the ~5.5 kbp *Stu*I-digested fragment from Ug8 yielded a
203 consensus sequence, the sequences obtained from the ~1.8 kbp fragment varied in length.

204 To resolve this problem and obtain the complete sequence, the RCA product derived from
205 Ug8 was digested with *Pst*I to yield a major fragment of ~7 kbp. The entire sequence of the
206 Ug8 isolate was subsequently obtained from alignment of sequences obtained from the
207 *Stu*I- and *Pst*I-derived fragments and from PCR-derived sequences used to confirm the
208 presence of the *Pst*I restriction site. Analysis of the RT/RNaseH-coding sequences of Ug8
209 revealed 92-98% similarity to BSUIV.

210

211 Sequence analysis

212 A summary of the characteristics of the complete genomes of each of the six BSV species is
213 presented in Table 3. The complete genomic sequences of BSUAV (sample Ke8, Genbank
214 accession number HQ593107), BSUIV (Ug8, accession number HQ593108, BSULV (Ug1,
215 accession number HQ593109), BSUMV (Ug12, accession number HQ593110), BSIMV (Ke10,
216 accession number HQ593112) and BSCAV (Ke171, accession number HQ593111) comprised
217 7519, 7458, 7401, 7532, 7769 and 7408 bp, respectively. For consistency with previous
218 conventions, numbering of each of the genomes begins with the 5' nucleotide of the
219 putative $tRNA^{met}$ priming site. Each of the six genomes contained three ORFs and, with the
220 exception of BSIMV ORF1, a conventional ATG initiation codon was present for all ORFs. A
221 non-conventional CTG initiation codon was predicted for BSIMV ORF1 based on sequence
222 comparisons with both BSMYV and BSVNV. The predicted size of proteins encoded by the
223 putative ORFs identified for each sequence is included in Table 3. The size of the intergenic
224 region in all six virus genomes ranged from 963 nt (BSUAV) to 1234 nt (BSIMV) and
225 contained a region with between 16-18 nucleotides complementary to the consensus
226 sequence of plant $tRNA^{met}$. Putative TATA boxes and polyadenylation signals, were also

227

Table 3 Genome features of East African BSV species

		ORF1			ORF2			ORF3			Transcriptional elements		
Badnavirus species	length total (nt)	length (nt)	start-stop (codon use)	protein size (kDa)	length (nt)	start-stop (codon use)	protein size (kDa)	length (nt)	start-stop (codon use)	protein size (kDa)	TATA	<gap>	polyA
Banana streak UA *virus* (BSUAV)	7519	534	483-1016 (atg-tga)	20.9	390	1013-1402 (atg-taa)	14.3	5637	1402-7038 (atg-taa)	216	ctcTATATAAgga	<56>	aataag
Banana streak UI *virus* (BSUIV)	7458	561	502-1062 (atg-tga)	21.6	336	1059-1394 (atg-taa)	12.3	5514	1395-6908 (atg-taa)	211	ctcTATATAAgga	<66>	aataaa
Banana streak UL *virus* (BSULV)	7401	561	532-1092 (atg-tga)	21.8	339	1089-1427 (atg-taa)	12.3	5502	1430-6931 (atg-tga)	211	ctcTATATAAgga	<64>	gataag
Banana streak UM *virus* (BSUMV)	7532	564	622-1185 (atg-tga)	21.7	312	1182-1493 (atg-tag)	11.7	5547	1497-7043 (atg-taa)	213	ggcTATATATAggt	<45>	aataaa
Banana streak IM *virus* (BSIMV)	7769	531	668-1198 (ctg-tga)	20.9	393	1195-1587 (atg-taa)	14.3	5613	1590-7202 (atg-taa)	215	atcTATAA--gag	<74>	aataaa
Banana streak CA *virus* (BSCAV)	7408	531	515-1045 (atg-tga)	21.1	405	1042-1446 (atg-taa)	14.7	5511	1446-6956 (atg-taa)	212	ctcTATAAATAgga	<55>	aataag

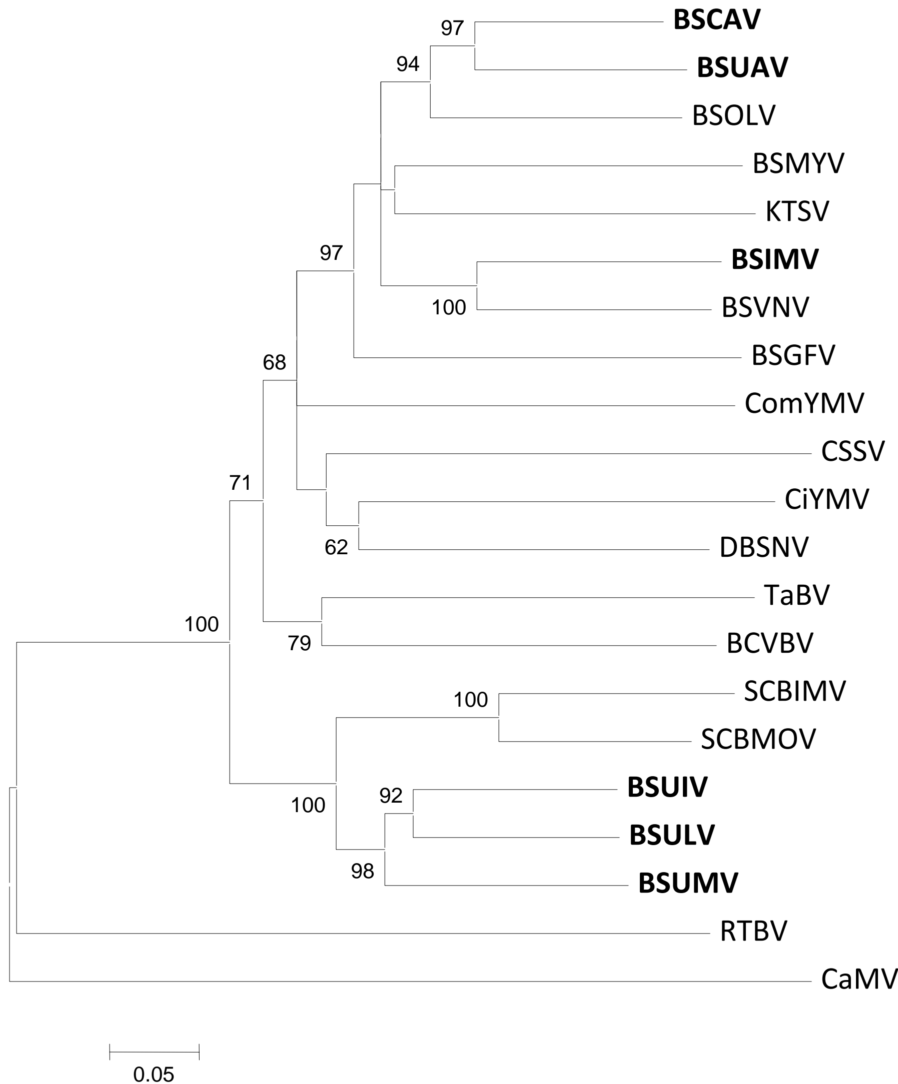
228 identified in the intergenic region of all sequences, 5’ of the putative $tRNA^{met}$ primer binding
229 site (Table 3). Analysis of the putative protein encoded by ORF 3 of each species revealed
230 the presence of several motifs that are highly conserved in badnavirus proteins including
231 movement, RNA-binding (zinc-finger motif), aspartyl proteinase, reverse transcriptase and
232 RNaseH.

233 Phylogenetic analysis of the RT/RNaseH region from full-length known episomal
234 badnavirus sequences showed that the three Ugandan BSVs (BSUIV, BSULV & BSUMV)
235 clustered together and that these viruses were more closely related to the sugarcane-
236 infecting badnaviruses than to other banana-infecting badnaviruses (Fig. 1). Two Kenyan
237 BSVs, BSUAV and BSCAV, were shown to be closely related and these formed a separate
238 cluster with BSOLV (Fig. 1). The Kenyan BSIMV did not cluster according to provenance but
239 instead was found to be most closely related to BSVNV. Pair-wise nucleotide similarities
240 within the RT/RNaseH-coding region of full-length sequences derived from the six BSV
241 sequences reported in this study together with their phylogenetically most closely-related
242 counterparts revealed at least a 20% nucleotide difference between the BSVs reported here
243 and other recognised badnaviruses (Table 4).

244

245 Detection of integrated sequences

246 To investigate the possible presence of integrated DNA of the six new BSV species in banana
247 genomic DNA, nucleic acid was extracted from a diverse collection of banana cultivars (Table
248 2) which had previously been certified as negative for episomal BSV sequences by ISEM and
249 RCA. The nucleic acid extracts were subsequently used in PCR with primers designed to



250

251 **Figure 1** Phylogenetic tree using neighbour-joining method (Kimura 2-parameter model with
252 bootstrapping (1000 replicates)) of the RT/RNaseH region of selected badnaviruses. *Rice tungro*
253 *bacilliform virus* (RTBV; genus *Tungrovirus*) and *Cauliflower mosaic virus* (CaMV; genus *Caulimovirus*)
254 were used as out-groups to the genus *Badnavirus*. GenBank accession numbers are: *Banana streak*
255 OL *virus* (BSOLV; GenBank accession NC_003381), *Banana streak* MY *virus* (BSMYV; GenBank
256 accession NC_006955), *Kalanchoe top-spotting virus* (KTSV; GenBank accession NC_004540), *Banana*
257 *streak* VN *virus* (BSVNV; Genbank accession AY750155), *Banana streak* GF *virus* (BSGFV; GenBank
258 accession NC_007002), *Commelina yellow mottle virus* (ComYMV; GenBank accession NC_001343),
259 *Cacao swollen shoot virus* (CSSV; GenBank accession NC_001574), *Citrus yellow mosaic virus* (CiYMV;
260 GenBank accession NC_003382), *Dioscorea bacilliform* SN *virus* (DBV; GenBank accession
261 DQ822073), *Taro bacilliform virus* (TaBV; Genbank accession AF357836), *Bougainvillea chlorotic vein*
262 *banding virus* (BCVBV; GenBank accession EU034539), *Sugarcane bacilliform* IM *virus* (SCBIMV;
263 GenBank accession NC_003031), *Sugarcane bacilliform* MO *virus* (SCBMOV; GenBank accession
264 NC_008017), RTBV (GenBank accession NC_001914), and CaMV (GenBank accession NC_001497).
265 Species whose genome was fully sequenced for the first time are shown in bold.

266

267

Table 4 Pair-wise distance matrix using core RT/RNaseH sequences of selected badnaviruses[a]

	BSCAV	BSUAV	BSOLV	BSGFV	BSUIV	BSULV	BSUMV	SCBIMV	SCBMOV	BSMYV	BSIMV	BSVNV	KTSV
BSCAV		23.2	26.1	38.3	45.8	43.1	47.1	50.8	47.9	38.9	33.5	34.8	37.4
BSUAV			30.6	40.0	45.6	46.3	48.6	51.2	52.8	37.3	35.7	34.4	37.0
BSOLV				43.4	44.4	48.2	46.4	53.6	50.8	38.0	28.0	38.2	34.6
BSGFV					50.3	47.8	47.9	58.1	53.9	45.2	42.1	40.2	41.9
BSUIV						23.2	27.9	38.1	34.2	47.8	52.7	49.9	50.4
BSULV							26.1	39.2	35.6	48.2	50.3	51.2	53.0
BSUMV								42.3	35.8	48.8	48.9	47.6	50.4
SCBIMV									25.0	59.0	57.4	60.3	60.0
SCBMOV										52.7	54.5	54.2	56.3
BSMYV											43.3	38.6	39.5
BSIMV												27.0	37.8
BSVNV													36.8
KTSV													

268 [a]Percent difference shown was calculated using the Kimura 2-parameter model following ClustalW alignment in MEGA 4.0.

269 specifically amplify a selected fragment of each of the six BSVs. Whereas no amplicons were
270 detected in extracts from any of the 12 banana cultivars tested for BSUAV, BSUIV, BSULV,
271 BSUMV or BSCAV, amplicons of the expected size were detected in extracts derived from
272 the four banana cultivars Ainu, FHIA-3, Balonkawe and Goly Goly Pot Pot tested for BSIMV
273 (Table 2). These results indicated that, for BSUAV, BSUIV, BSULV, BSUMV or BSCAV, the
274 sequences delimited by these primers were not present as either episomal or integrated
275 DNA in any of the 12 cultivars tested. In contrast, the genomic DNA of four of the 12
276 cultivars tested appeared to contain an integrated BSIMV sequence.

277

278 **DISCUSSION**

279 We have completely sequenced the genomes of six BSV isolates collected in East Africa, for
280 which only partial sequences have been previously available. All six sequences had a typical
281 badnavirus genome organisation and contained the conserved motifs characteristically
282 found in the putative ORF 3 polyprotein of badnaviruses. Based on the criteria for
283 recognition of distinct species in the genus *Badnavirus* (a difference in the nucleotide
284 sequence of the RT/RNaseH-coding region of more than 20% (Hull *et al*., 2005)), we propose
285 that the six isolates should be recognised as new BSV species and be designated as BSUAV,
286 BSUIV, BSULV, BSUMV, BSIMV and BSCAV based on a recent amendment to the naming
287 convention for badnaviruses (Geering, 2010). The results from this study confirm the
288 presence, in Uganda, of three of the 13 putative BSV species reported previously (Harper *et*
289 *al*., 2005). Further, this is the first report of BSCAV from Kenya, and confirms a previous
290 report of BSUAV and BSIMV in Kenya using IC-PCR (Karanja *et al*., 2008). Importantly, the
291 detection of BSUAV, BSUIV, BSULV and BSUMV in this study is based on episomal DNA

292 amplified by RCA, and not IC-PCR which may detect integrated sequences (Le Provost *et al*.,
293 2006, Iskra-Caruana *et al*., 2009), as in previous studies.

294 Previous phylogenetic analyses of full-length and partial sequences have consistently
295 identified three distinct clades of banana-infecting badnaviruses (Harper *et al*., 2005;
296 Bousalem *et al*., 2008; Gayral and Iskra-Caruana, 2009). The isolates reported in this study
297 grouped within two of the three clades, consistent with previous reports. BSUIV, BSULV and
298 BSUMV grouped within clade 3 which also includes badnavirus species characterised from
299 sugarcane while BSUAV, BSCAV and BSIMV grouped within clade 1, which contained only
300 badnavirus species originating from banana. Interestingly, however, the sequence of our
301 BSCAV isolate also showed 93-96% homology at the nucleotide level to six unpublished
302 *Sugarcane bacilliform virus* (ScBV) isolates in the GenBank database. As such, ScBV isolates
303 now appear to group within two different clades (clades 1 and 3), suggesting that the
304 movement of badnaviruses across the host-plant boundary between sugarcane and banana
305 has likely occurred on more than one occasion.

306 Several authors (Bousalem *et al*., 2008; Gayral and Iskra-Caruana, 2009) have
307 reported a close phylogenetic relationship between BSULV and the proposed BSUgKV
308 identified by Harper *et al* (2005). A comparison of the RT/RNaseH-coding sequence of our
309 BSULV isolate to other badnavirus sequences revealed 82% homology to four sequences
310 described as BSUgKV (Harper *et al*., 2005). This finding suggests that BSULV and BSUgKV are
311 a single species.

312 The sequences of several BSV species are known to occur within the *M. balbisiana*
313 genome, the presence of which has provided a challenge for the diagnosis of episomal BSV
314 infection using PCR-based methodologies. As a preliminary study to determine whether

315 sequences related to the six viruses reported in this study had integrated counterparts, 12
316 banana genotypes known to be free of episomal BSV were tested for each of the six viruses
317 by PCR using virus-specific primers. None of the samples tested positive for BSUAV, BSUIV,
318 BSULV, BSUMV or BSCAV, while four plants, all containing B-genomes, tested positive for
319 BSIMV. The detection of integrated BSIMV sequences in banana accessions with a *M.*
320 *balbisiana* genome component is consistent with previous reports (Geering *et al*., 2005A,
321 Gayral *et al*., 2010). Further, when the sequences of the core RT/RNaseH-coding region of
322 the six BSV species described in this manuscript were used to search for homologous
323 endogenous sequences (both endogenous badnavirus and *Musa* genomic BAC sequences) in
324 GenBank, BSIMV showed 99% homology to several endogenous badnavirus sequences and
325 94% homology to a *Musa balbisiana* BAC sequence (GenBank accession AP009334). In
326 contrast, the remaining five species did not produce a significant match (i.e. 80% similarity
327 or greater) to either endogenous badnavirus or *Musa* BAC sequences. While our findings do
328 not conclusively exclude the presence of integrated sequences for each of the other five BSV
329 species in the genotypes tested, the results suggest that PCR might be a suitable tool for
330 diagnosis of these species in bananas with selected genetic backgrounds.

331 Diagnosis of BSV using the non-sequence specific RCA assay should dramatically
332 improve the scope of detection of heterogeneous mixture of viruses comprising the BSV
333 complex. Further, the full-length sequences presented here will improve the opportunity to
334 diagnose BSV infections using restriction-digest based RCA assays. Additional work is still
335 required, however, to confirm the episomal nature of the additional Ugandan BSV isolates
336 reported by Harper *et al*., (2005). Biological information pertaining to the BSV species

337 reported in this work will be useful to understand the BSV-banana system with more clarity,
338 and allow the development of improved diagnostic tests.

339

340 **ACKNOWLEDGEMENTS**

341 This research was funded by the Bill and Melinda Gates Foundation Grand Challenges in
342 Global Health Program. We are grateful to our colleagues Dr. Charles Changa and Dr.
343 Jerome Kubiriba at the National Agricultural Research Organisation (NARO), Uganda, as well
344 as Dr. Laura Karanja at the Kenyan Agricultural Research Institute (KARI), Njoro, Kenya, for
345 assisting with field collections and leaf samples of banana.

346

347 **REFERENCES**

348 Bousalem, M., Douzery, E.J.P., Seal, S.E. (2008) Taxonomy, molecular phylogeny and
349 evolution of plant reverse transcribing viruses (family *Caulimoviridae*) inferred from
350 full-length genome and reverse transcriptase sequences. *Archives of Virology*, **153**,
351 1085-1102.

352 Cheng, C.P., Lockhart, B.E.L., Olszewski, N.E. (1996) The ORFI and II proteins of Commelina
353 yellow mottle virus are virion-associated. *Journal of General Virology*, **223**, 263-271.

354 Côte, F.X., Galzi, S., Folliot, M., Lamagnére, Y., Teycheney, P-Y., Iskra-Caruana, M-L. (2010)
355 Micropropagation by tissue culture triggers differential expression of infectious
356 endogenous Banana streak virus sequences (eBSV) present in the B genome of natural
357 and synthetic interspecific banana plantains. *Molecular Plant Pathology,* **11,** 137-144.

358 Dahal, G., Hughes, J., Gauhl, F., Pasberg-Gauhl, C., Nokoe, K.S. (2000) Symptomatology and
359 development of banana streak, a disease caused by banana streak badnavirus, under
360 natural conditions in Ibadan, Nigeria. *Acta Horticulturae (ISHS)*, **540**, 361-375.

361 Dallot, S., Acuña, P., Rivera, C., Ramírez, P., Côte, F., Lockhart, B.E.L., Caruana, M.L. (2001)
362 Evidence that the proliferation stage of micropropagation procedure is determinant in
363 the expression of banana streak virus integrated into the genome of the FHIA 21
364 hybrid (Musa AAAB). *Archives of Virology*, **146**, 2179-2190.

365 Gayral, P., Blondin, L., Guidolin, O., Carreel, F., Hippolyte, I., Perrier, X., Iskra-Caruana, M-L.,
366 (2010) Evolution of endogenous sequences of Banana streak virus: What can we learn
367 from banana (Musa sp.) evolution? *Journal of Virology*, **84**, 7346 – 7359.

368 Gayral, P., Iskra-Caruana, M.L. (2009) Phylogeny of Banana streak virus reveals recent and
369 repetitive endogenization in the genome of its banana host (Musa sp.). *Journal of*
370 *Molecular Evolution*, **69**, 65-80.

371 Gayral, P., Noa-Carrazana, J-C., Lescot, M., Lheureux, F., Lockhart, B.E.L., Matsumoto, T.,
372 Piffanelli, P., Iskra-Caruana, M-L. (2008) A single Banana streak virus integration event
373 in the banana genome as the origin of infectious endogenous pararetrovirus. *Journal*
374 *of Virology*, **82**, 6697-6710.

375 Geering, A.D.W. (2009) The distribution of badnavirus-infected Musa germplasm from the
376 international transit centre, Katholieke Universiteit Leuven, Belgium. In *International*
377 *Symposium on Recent Advances in Banana Crop Protection for Sustainable Production*
378 *and Improved Livelihoods*, pp85-88. Eds D. Jones and I. Van den Bergh. ISHS.

379 Geering, A.D.W. (2010) Five new species and four nomenclature changes in the genus
380 Badnavirus. [online] ICTV Taxonomy Proposals - Plant. Available at
381 http://talk.ictvonline.org/files/proposals/taxonomy_proposals_plant1/m/plant01/221
382 6.asp [Accessed 14 December 2010].

383 Geering, A.W.D., McMichael, L.A., Dietzgen, R.G., Thomas, J.E. (2000) Genetic diversity
384 among banana streak virus isolates from Australia. *Phytopathology*, **90**, 921-927.

385 Geering, A.D.W., Olszewski, N.E., Harper, G., Lockhart, B.E.L., Hull, R., Thomas, J.E. (2005A)
386 Banana contains a diverse array of endogenous badnaviruses. *Journal of General*
387 *Virology*, **86**, 511-520.

388 Geering, A. D. W., Pooggin, M.M., Olszewski, N.E., Lockhart, B.E.L., Thomas, J.E. (2005B)
389 Characterisation of banana streak Mysore virus and evidence that its DNA is
390 integrated in the B genome of cultivated Musa. *Archives of Virology*, **150**, 787-796.

391 Geering, A.W.D., Scharaschkin, T., Teycheney, P.Y. (2010) The classification and
392 nomenclature of endogenous viruses of the family Caulimoviridae. *Archives of*
393 *Virology*, **155**, 123-131.

394 Harper, G., Hart, D., Moult, S., Hull, R. (2004) Banana streak virus is very diverse in Uganda.
395 *Virus Research*, **100**, 51-56.

396 Harper, G., Hart, D., Moult, S., Hull, R., Geering, A., Thomas, J. (2005) The diversity of banana
397 streak virus isolates in Uganda. *Archives of Virology*, **150**, 2407-2420.

398 Harper, G., Osuji, J.O., Heslop-Harrison, J.S., Hull, R. (1999) Integration of banana streak
399 badnavirus into the Musa genome: molecular and cytogenetic evidence. *Virology*, **255**,
400 207-213.

401 Hull, R., Geering, A., Harper, G., Lockhart, B.E., Schoelz, J.E. (2005) Family Caulimoviridae. In
402 *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*
403 p. 385-396. Eds C.M. Fauquet, M.A. Mayo, J. Maniloff, U. Desselberger and L.A. Ball.
404 London: Elsevier.

405 Iskra-Caruana, M.L., Gayral, P., Galzi, S., Laboureau, N. (2009) How to control and prevent
406 the spread of banana streak disease when the origin could be viral sequences
407 integrated in the banana genome? In *International Symposium on Recent Advances in*
408 *Banana Crop Protection for Sustainable Production and Improved Livelihoods*, pp77-84.
409 Eds D. Jones and I. Van den Bergh. ISHS.

410 Jacquot, E., Hagen, L.S., Jacquemond, M., Yot, P. (1996) The open reading frame 2 product
411 of Cacao swollen shoot badnavirus is a nucleic acid-binding protein. *Virology*, **225**,
412 191-195.

413 James, A.P., Geijskes, R.J., Dale, J.L., Harding, R.M. (2011) Development of a novel rolling-
414 circle amplification technique to detect banana streak virus which also discriminates
415 between integrated and episomal virus sequences. *Plant Disease*, **95**, 57-62.

416 Karanja, L., Wangai, A., Harper, G., Pathak, R.S. (2008) Molecular identification of Banana
417 streak virus isolates in Kenya. *Journal of Phytopathology*, **156**, 678-686.

418 Kimura, M. (1980) A simple method for estimating evolutionary rate of base substitutions
419 through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*,
420 **16**, 111-120.

421 Le Provost, G., Iskra-Caruana, M-L., Acina, I., Teycheney, P-Y. (2006) Improved detection of
422 episomal Banana streak viruses by multiplex immunocapture PCR. *Journal of*
423 *Virological Methods*, **137,** 7–13.

424 Leclerc, D., Burri, L., Kajava, A.V., Mougeot, J.L., Hess, D., Lustig, A., Kleeman, G., Hohn, T.
425 (1998) The open reading frame III product of Cauliflower mosaic virus forms a
426 tetramer through a N-terminal coiled-coil. *The Journal of Biological Chemistry*, **237**,
427 29015-29021.

428 Lheureux, F., Laboureau, N., Muller, E., Lockhart, B.E.L., Iskra-Caruana, M.L. (2007)
429 Molecular characterisation of banana streak acuminata Vietnam virus isolated from
430 Musa acuminata siamea (banana cultivar). *Archives of Virology*, **152**, 1409-1416.

431 Ndowora, T., Dahal, G., LaFleur, D., Harper, H., Hull, R., Olszewski, N.E., Lockhart, B. (1999)
432 Evidence that badnavirus infection in Musa can originate from integrated
433 pararetroviral sequences. *Virology*, **255**, 214-220.

434 Tamura, K., Dudley, J., Nei, M., Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics
435 Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596-
436 1599.

437 Yang, I.C., Hafner, G.J., Revill, P.A., Dale, J.L., and Harding, R.M. (2003) Sequence diversity
438 of South Pacific isolates of taro bacilliform virus and the development of a PCR-based
439 diagnostic test. *Archives of Virology*, **148**, 1957–1968.