Available online 20 January 2014; Published in VIRUS RESEARCH, 24 June 2014.

The prevalence of badnaviruses in West African yams (*Dioscorea* spp.) and evidence of endogenous pararetrovirus sequences in their genomes.

Susan Seal[a,b], Aliyu Turaki[a], Emmanuelle Muller[b], P. Lava Kumar[c], Lawrence Kenyon[a, d], Denis Filloux[b], Serge Galzi[b], Antonio Lopez-Montes[c], Marie-Line Iskra-Caruana[b]

[a]*Natural Resources Institute, University of Greenwich, Central Avenue, Chatham Maritime, Kent ME4 4TB, UK;*

[b]*CIRAD, UMR BGPI, F-34098 MONTPELLIER, France*

[c]*International Institute of Tropical Agriculture (IITA), Oyo Road PMB 5320, Ibadan, Nigeria*

[d]*Present address: AVRDC, PO Box 42, Shanhua, Tainan 74199, Taiwan*

Corresponding author Tel: +44 1634 880088, Fax: +44 1634 883379

E-mail address: s.e.seal@gre.ac.uk

ABSTRACT

Yam (*Dioscorea* spp.) is an important vegetatively-propagated staple crop in West Africa. Viruses are pervasive in yam worldwide, decreasing growth and yield, as well as hindering the international movement of germplasm. Badnaviruses have been reported to be the most prevalent in yam, and genomes of some other badnaviruses are known to be integrated in their host plant species. However, it was not clear if a similar scenario occurs in *Dioscorea* yam. This study was conducted to verify the prevalence of badnaviruses, and determine if badnavirus genomes are integrated in the yam genome.

Leaf samples (n=58) representing eight species of yam from global yam collections kept at CIRAD, France, and 127 samples of *D. rotundata* breeding lines (n=112) and landraces (n=15) at IITA, Nigeria, were screened using generic badnavirus PCR primers. Positive amplification of an expected ca. 579 bp fragment, corresponding to a partial RT-RNaseH region, was detected in 47 (81%) of 58 samples analyzed from CIRAD collections, and 100% of the 127 IITA *D. rotundata* samples. All the *D. cayenensis* and *D. rotundata* samples from the CIRAD and IITA collections tested PCR-positive, and sequencing of a selection of the PCR products confirmed they were typical of the genus *Badnavirus.* A comparison of serological and nucleic acid techniques was used to investigate whether the PCR-positives were sequences amplified from badnavirus particles or putative endogenous badnavirus sequences in the yam genome. Protein A sandwich-enzyme-linked immunosorbent assay (PAS-ELISA) with badnavirus polyclonal antisera detected cross-reacting viral particles in only 60% (92 of 153) of the CIRAD collection samples analyzed, in contrast to the aforementioned 81% by PCR. Immunosorbent electron microscopy (ISEM) of virus preparations of a select set of 16 samples, representing different combinations of positive and negative PCR and PAS-ELISA results, identified bacilliform particles in 11 of these samples. Three PCR-positive yam samples from Burkina Faso (cv. Pilimpikou) were identified in which no viral particles were detected by either PAS-ELISA or ISEM. Southern hybridisation results using a yam badnavirus RT-RNaseH sequence (Gn155Dr) as probe, supported a lack of badnavirus particles in the cv. Pilimpikou and identified their equivalent sequences to be of plant genome origin. Probe Gn155Dr, however, hybridised to viral particles and plant genomic DNA in three *D. rotundata* samples from Guinea. These results represent the first data demonstrating the presence of integrated sequences of badnaviruses in yam. The implications of this for virus-indexing, and breeding and multiplication of seed yams are discussed.

Abbreviations

DaBV = *Dioscorea alata* bacilliform virus, DbBV = *Dioscorea bulbifera* bacilliform virus, DBALV = *Dioscorea bacilliform alata virus,* DBSNV = *Dioscorea bacilliform sansibarensis virus*, DBV = *Dioscorea* bacilliform virus, eDBV = endogenous *Dioscorea* bacilliform virus, EPRV = endogenous pararetrovirus, PVCV= *Petunia vein-clearing virus*, TVCV = *Tobacco vein-clearing virus.*

Keywords:

Badnavirus

Dioscorea yam

ELISA

Endogenous pararetroviruses

Immunosorbent electron microscopy

PCR

1. Introduction

Yam *(Dioscorea* species) is the fourth most important food tuber crop in the world after potato, sweet potato, and cassava (FAO, 2012). In West Africa, it is the second most important food crop after cassava by value and production (FAO, 2012; Scarcelli et al., 2006). It plays an essential role in food security and income generation for smallholders, particularly in West Africa which produces about 95% of the world's total yam production (Asiedu and Sartie, 2010; IITA, 2012; Mignouna et al., 2008). The white Guinea yam, *Dioscorea rotundata,* is the predominant yam species grown in this region, in contrast to other popular yam species such as *D. alata* and *D. esculenta* which predominate in the South Pacific (Kenyon et al., 2008). The yellow Guinea yam, *D. cayenensis* is also cultivated widely in West Africa, and like *D. rotundata*, is an African domesticated species originating from wild Dioscoreaceae of the Enantiophyllum Uline section. It should be noted that the classification of Guinea yams into either *D. rotundata* Poir. or *D. cayenensis* Lam. has, however, been confused for a long time, and they have in the past few decades often been referred to collectively as members of the *D. cayenensis-rotundata* species complex (Dumont et al., 2006).

Yam is generally propagated vegetatively through its tubers. This facilitates the accumulation of pathogens, particularly viruses of which there are at least 26 different species belonging to nine virus genera reported in yams worldwide to date (Bousalem et al., 2009; Kenyon et al., 2001). These virus infections have the potential to reduce tuber yields and quality, and impede yam germplasm movement and thus hinder international exchange of selected improved varieties (Bousalem et al., 2009; Kenyon et al., 2008). The scarcity and associated high expense of 'clean seed' yam has been identified as one of the most important critical constraints to increasing yam production and productivity in West Africa (IITA, 2012).

Several surveys on yam viruses suggest that badnaviruses are the most prevalent globally (Bousalem et al., 2009; Eni et al., 2008a, b; 2009; Galzi et al., 2013; Kenyon et al., 2008). Badnavirus particles were first reported in yam in association with a flexuous virus, causing internal brown spot disease in *D. alata* and *D. cayenensis* in the Caribbean (Harrison and Roberts, 1973; Mantell and Haque, 1978). Two decades later yam badnaviruses were characterised by their nucleic acid and serological properties; particles isolated from *D. alata* and *D. bulbifera* were partially characterised and named informally as Dioscorea alata bacilliform virus (DaBV) and Dioscorea bulbifera bacilliform virus (DbBV) (Briddon et al., 1999; Phillips et al., 1999). These viruses were reported to induce leaf distortions and veinal chlorosis (Phillips et al., 1999), although others found that often infected plants show no marked symptoms (Kenyon et al., 2008; Seal and Muller, 2007).

Current taxonomic criteria (King et al., 2012) recognise only two species of yam badnavirus, for which complete genome sequence data (ca. 7.2-7.4 kb) exist, namely *Dioscorea bacilliform alata*

virus (DBALV) isolated from *D. alata* in Nigeria (Briddon et al., 1999), and *Dioscorea bacilliform sansibarensis virus* (DBSNV) present in a wild *D. sansibarensis* from Benin (Seal and Muller, 2007). At least a further 10 putative *Badnavirus* species are indicated to be present in *Dioscorea* species globally through possessing partial (529 bp) RT-RNaseH nucleotide sequences that differ by more than the International Committee on Taxonomy of Viruses (ICTV) recommended species demarcation threshold for this region of >20% (Bousalem et al., 2009; King et al., 2012).

Sensitive virus diagnostic tests are required to enable the identification of virus-free seed yams, and will underpin current efforts in West Africa to generate and multiply disease-free yam planting material (IITA, 2012), as well as being essential to generate meaningful data in field surveys and epidemiology. Of the three virus genera (*Badnavirus*, *Cucumovirus* and *Potyvirus*) known to be of economic importance to yams in West Africa, reliable diagnostic tests exist for detection of yam potyviruses and cucumoviruses (Eni et al., 2008b; 2009; Mumford and Seal, 1997; Wylie et al., 1993). The serological and genetic heterogeneity of yam badnaviruses, however, poses a challenge for the development of diagnostic tests, as also experienced for badnaviruses in a wide range of other crops (Harper et al., 2005; Kenyon et al., 2008; Lockhart, 1986; Muller et al., 2011). Furthermore, the discovery of DNA sequences of the genus *Badnavirus* as integrated sequences in their plant host genome complicates the use of nucleic-acid based diagnostics, as illustrated by the challenges experienced in reliable detection of virus particles of banana streak viruses (BSVs) in *Musa* species (Harper et al., 1999b; Ndowora et al., 1999; Le Provost et al., 2006). Such integrated sequences appear to be a common phenomenon within genera of the family *Caulimoviridae*, and are termed endogenous pararetroviruses (EPRVs) (Geering et al., 2010; Mette et al., 2002; Staginnus et al., 2009).

The structure of EPRV sequences can be complex, and generally consists of rearranged patterns showing tandem repeats, fragmentations, inversions and duplications of the viral genome or parts thereof (Chabannes et al., 2013; Gayral et al., 2008; Ndowora et al., 1999; Richert-Pöggeler et al., 2003). Although most EPRVs reported seem to be simply neutral components in their host plant genomes, there have been three host examples to date which are of concern to breeding and virus-indexing programmes as they are 'activatable', i.e. episomal virus infections can be initiated *de novo* from these sequences integrated in their host plant genomes (Chabannes et al., 2013; Lockhart *et al*., 2000; Richert-Pöggeler et al., 2003). The activatable EPRVs represent three species of the genus *Badnavirus* discovered in banana genomes of *Musa balbisiana* species, namely *Banana streak OL virus* (BSOLV), *Banana streak Imové virus* (BSImV), and *Banana streak GF virus* (BSGFV) (Chabannes et al., 2013; Gayral et al., 2008; Harper et al., 1999a; Iskra-Caruana et al., this volume; Ndowora et al., 1999), as well as the petuvirus *Petunia vein clearing virus* (PVCV) in petunia (Richert-Pöggeler et

al., 2003), and solendovirus *Tobacco vein-clearing virus* (TVCV) in tobacco (Jakowitsch et al., 1999; Lockhart *et al*., 2000). Activation is considered in banana to be triggered by the epigenetic modifications that occur during hybridization of parental genomes as well as environmental stresses (e.g. wounding, tissue culture, and drought) (Dallot et al., 2001; Cote et al., 2010; Harper et al., 2002). Episomal virus has been suggested to be generated and released from EPRV sequences through mechanisms involving either homologous recombination between repeat regions, and/or by direct reverse transcription (Chabannes and Iskra-Caruana, 2013; Harper et al., 2002; Iskra-Caruana et al., 2010, this volume; Ndowora et al., 1999; Richert-Pöggeler et al., 2003).

Previous studies on yam badnaviruses have reported unusually high levels (91-96%) of badnavirus PCR-positive samples within *Dioscorea rotundata* collections from both Benin and Guadeloupe (Bousalem et al., 2009), in comparison to much lower ELISA-positives (26-35%) from South Pacific yam samples (Kenyon et al., 2008). This study was initiated to investigate whether the unusually high PCR-positive results in *D. rotundata* samples represented virus particle infections, or might be the result of PCR amplification of previously unidentified endogenous Dioscorea bacilliform virus sequences (termed eDBVs, according to the nomenclature proposed by Staginnus et al., 2009) in this yam species genome. Data revealed some samples from the *D. cayenensis-rotundata* complex to contain eDBVs. A high prevalence of serologically and genetically diverse badnaviruses was also detected in the West African yam collections. These findings have implications for the generation of high quality breeder and foundation yam planting material, and for the maintenance of its virus-free status when cultivated in the field.

2. Material and methods

2.1 Plant samples

Yams (n = 153) from the CIRAD-IRD collections were maintained and leaves collected from glasshouses as described previously (Bousalem et al., 2009). They were of eight species: *D. abyssinica, D. alata, D. cayenensis, D. dumetorum, D. nummularia, D. rotundata, D. sansibarensis and D. trifida*. The IITA collection samples (n=127) consisted of first filial (F^1) generations of breeding lines of *D. rotundata* (n=112) and *D. rotundata landraces* (n=15), which were collected from screen houses at the International Institute of Tropical Agriculture (IITA, Ibadan Nigeria) and placed in polythene bags (14×14 cm, Polybags Ltd., UK). Details of all samples from these collections are given in the Supplementary materials. It should be noted that the yams from Guinea, Benin and Burkina Faso have all been classified as *D. rotundata,* but in fact come from a range of varieties, the genetic relationships of which are not well understood. All that is currently known is that not all of these samples are *D. rotundata* sensu stricto, and should currently be considered simply as

cultivated yams, of African origin, belonging to the Enantiophyllum section (Roland Dumont and Philippe Vernier, personal communication). The landrace cv. Pilimpikou (also known as 'Bolgo Nyu') yam, are the most genetically distinct from this group, and are grown only in a small central area of Burkina Faso, about 100 km north of Ouagadougou (Goudou-Urbino et al., 1996). They are male yams and are neither *D. cayenensis* nor *D. rotundata* sensu stricto (Dumont et al., 2006). As a precise classification is not available for the '*D. rotundata*' samples in which the eDBV has been identified, their genomes have been described collectively here as *D. cayenensis-rotundata*.

Symptoms shown by the samples varied greatly, from no obvious visible symptoms to veinal chlorosis and leaf distortions (Fig.1).

2.2 Total DNA extractions from yam leaves and PCR amplification of badnavirus genomic segments

A modification of the Lodhi et al. (1994) method was used as described previously (Kenyon et al., 2008). Selected leaf samples (n=58) were ground within polythene disposable bags to reduce the possibility of contamination between samples**.** The extracted DNAs were purified through Tip20 columns (Qiagen) according to manufacturer's instructions and resuspended in 100 µl of sterile distilled deionised water (SDW). Total DNAs were screened for the presence of badnavirus sequences using the generic badnavirus primer pair Badna-FP/-RP (Yang et al., 2003), which amplify a ~579 bp region (529 bp excluding primer sequences) of the of RT-RNaseH region, which has been used for taxonomic purposes within this genus (King et al., 2012).

2.3 Protein A sandwich (PAS)-ELISA procedure

Samples were tested by PAS-ELISA using two antisera developed to yam badnaviruses , as well as a polyclonal mix that cross reacts with badnaviruses in general. Yam badnavirus antisera 'DaBV' (batch '500') and 'DbBV' (batch '428') (Phillips et al., 1999) were kindly provided by Dr Nicola Spence (UK), whereas the general badnavirus polyclonal mix 'BenL' was provided by Prof. Ben Lockhart (USA). Yam leaves ground directly in individual polythene bags in (1/10, w/v) 1x PBS-T containing 2% (w/v) polyvinylpyrrolidone-40 (PVP-40) and 1% (w/v) Na_2SO_3, as described by Mumford and Seal (1997). Known badnavirus-free and virus-infected yam leaves were included as negative and positive controls, respectively. Samples and positive controls were loaded in duplicate whereas four wells were used for the negative control. The outer rows of wells were not used. A standard PAS-ELISA protocol was used and specific details are as described in Kenyon et al., (2008), using trapping and detection dilutions of 1:200 (v/v) for BenL antisera, and 1:2000 (v/v) for DaBV and DbBV antisera. Para-nitrophenyl diphosphate sodium salt was used as the colorimetric reagent to measure the amount of bound alkaline phosphatase-labelled antibody in the final detection step.

Absorbances were recorded at 405 nm after 1 h and 18 h incubation at 37°C. The means of duplicate samples having an absorbance lower than the following equation: '(means of healthy) + (3x standard deviation of the healthy)' were assigned negative (-). Otherwise, they were scored positive with (+) representing samples with an absorbance one time higher than the value of the threshold, (++) when twice higher, and (+++) when three times higher.

2.4 Virus semi-purification preparation for immunosorbent electron microscopy (ISEM)

Semi-purifications of badnavirus particles were prepared essentially as described by Lockhart, (1986): 18 ml 0.5M Na-K phosphate buffer pH7.4 (2 volumes K_2HPO_4 buffer to 1 volume NaH_2PO_4) containing polyvinylpyrrolidone (average molecular weight 40,000) (40 g/l), urea (40 g/l) and B-mercaptoethanol (5 ml/l) were added to yam leaf material (10 g) ground in pestle and mortars using liquid nitrogen. The suspension was filtered through Miracloth, 1 ml 33% Triton X100 added, and the mixture shaken 15 min at room temperature, followed by centrifugation (14000 g, 10 min). Supernatants were transferred into ultracentrifuge tubes, and a cushion of 6 ml 30% (w/v) saccharose in 0.1M pH7.4 Na+K phosphate buffer added. Following centrifugation (35000 rpm = 148000 *g* max in Beckmann LE80K centrifuge with 50.2 Ti rotor) for 90 min, supernatants were discarded and tubes rinsed with SDW to eliminate detergent residues. Pellets were resuspended in 100 µl 0.015M Tris-HCl pH7.4 buffer, centrifuged (14000 g, 15 min) and supernatant recovered for storage at -20°C. Extracts were examined by ISEM on Ben-L antisera-coated carbon grids, using 10 µl extract per grid.

2.5 Southern blotting and hybridisation protocol

Approximately 20-25 µg of total yam DNAs were digested overnight with at least 50 U (units) of restriction enzymes *Sph*I, *Nde*I and *Eco*RV at the manufacturer's recommended reaction conditions. These restriction enzymes were selected by determining from the full length sequence information of DBALV (Briddon et al., 1999) those which may cut the viral genome just once, as this would best differentiate the linearised viral bands from plant genomic bands. Multiple cutters were not appropriate as these would be less likely to reveal integrated sequences through the presence of hybridising bands >8 kb. An additional technical challenge faced was to obtain sufficient quantities (>10 ug per gel lane) of yam total DNA preparations of sufficient purity to enable complete restriction enzyme digestion.

Digested DNA was separated by 0.8% (w/v) agarose gel electrophoresis, after which DNAs in gels were transferred by capillary transfer to a Hybond N+ membrane (GE Healthcare, UK) according to manufacturer's instructions using the alkaline transfer protocol. Membranes were baked at 80 °C

for 2 h to fix DNA to membranes, and then stored dry till use. Probes were prepared by labelling gel-purified PCR-amplified RT-RNaseH fragment inserts of clones with ^{32}P-dCTP according to the Megaprime DNA labelling protocol (GE Healthcare, UK). Labelling was allowed to proceed for 30 min, and then unincorporated nucleotides were removed using a Nick column (Pharmacia, UK) to lower non-specific background readings from the probe. Hybridisation was done using 15 ml (per 400 cm^2 membrane) consisting of 3.9 ml SSC-Dextran, 1.4 ml 20x SSC, 1.4 ml Denhardt's solution 50x, 0.35 ml 20% w/v SDS, 0.1 ml 10 mg/ml denatured herring sperm DNA and 7.85 ml sterile distilled deionised water. Membranes were transferred to this mix whilst probes and herring sperm DNA were being denatured at 95^{o}C for 5 min. The denatured probes were snap-cooled on ice, or added directly to the membranes. Hybridisation was allowed to proceed overnight at 65^{o}C. The membranes were then washed twice at 65^{o}C in 2x SSC, 0.1% SDS, twice in 0.4x SSC, 0.1% SDS, ending with a 1 h wash at 60^{o}C in 0.2x SSC, 0.1% SDS.

2.6 Sequence analysis

A phylogenetic tree was generated from sequences of samples included in the Southern hybridisation studies. The sequences obtained were aligned using the Clustal W multiple alignment algorithm in the Bioedit package (http://www.mbio.ncsu.edu/bioedit/bioedit.html). The phylogenetic tree was created by the Neighbour-Joining method using the Darwin 5 software (Perrier et al., 2003). The robustness of the tree generated was determined by bootstrap sampling of the multiple alignment (1,000 replicates) and the cut-off value was 70%. Isolates with more than 20% nucleotide sequence divergence within the RT-RNaseH-coding region are considered to be separate badnavirus species according to ICTV guidelines (King et al., 2012). The sequence comparison of the RT-RNaseH coding region in this study was based on the 529 bp sequence internal to the Badna-FP and Badna-RP primers reported by Yang et al., (2003).

3. Results

3.1 PCR amplification of badnavirus sequences

Forty seven of the 58 CIRAD samples (81%) tested were scored as PCR-positive (Fig. 2, Table 1, and Supplementary materials). All 35 samples tested from the collections from West Africa (10 from Benin, 14 from Guinea, and 11 from IRD-Africa) were PCR-positive, in contrast to only 9 of 14 (64%), and 3 of 9 (33%) yam samples from Central and Southern America and the South Pacific (Vanuatu) respectively. A striking difference between the composition of these collections is that 29 of the 35 samples from the African collections were of *D. cayenensis-rotundata*, whereas there were

none of this species complex represented in either the Central America or Vanuatu collections. Every *D. cayenensis-rotundata* (n=29) plant tested from the collections was PCR-positive. Screening of the IITA-maintained *D. rotundata* breeding lines (n=112) and landraces (n = 15), also generated 100% badnavirus PCR-positive results for these 127 samples (Fig. 3, details of samples given in Supplementary materials). Samples tested of *D. dumetorum* (n=2), *D. esculenta* (n=1), D. *sansibarensis* (n=1) and *D. trifida* (n=4) were also 100% positive, but much lower sample numbers were tested of these species. For the remaining two species tested by PCR, about half the samples tested were positive (7 of 15 *D. alata*, and 2 of 4 *D. nummularia,* Table 1).

Sequencing of a selection of the bands identified as representing PCR-positive results confirmed that these products were typical of members of the genus *Badnavirus.* Phylogenetic analyses of these sequences have been published previously (Bousalem et al. 2009) and revealed the RT-RNaseH sequences (n=47) from the CIRAD-IRD yam collections to fall into five of the 12 yam badnavirus putative species groups, namely Dioscorea bacilliform virus (DBV)-A (containing DBALV), DBV-B, DBV-C, DBV-D, and DsBV. The latter species group has since been named as DBSNV (King et al., 2012). The Bousalem et al. (2009) terminology for the species groups has been used in this study rather than the yam badnavirus groupings (Group 1-11) of Kenyon et al. (2008), and the correspondence of these two different classification systems is shown in Fig.4.

Sequences generated from West African samples of the *D. cayenensis-rotundata* complex screened by PCR and ELISA in this study, contained sequences belonging to only three of these groups (DBV-A, DBV-B, and DBV-D) (Table 2). The phylogenetic relationships of the nucleotide sequences originating from material used in the Southern hybridisation studies are shown in Fig. 4. It should be noted that the sequencing carried out on samples will only have revealed a proportion of the total diversity of badnavirus RT-RNaseH sequences present in these samples.

3.2 PAS-ELISA detection of badnavirus particles

To determine whether virus particles occurred in samples possessing these different *Badnavirus* species sequence groups, samples (n=153) in the CIRAD yam collection were screened by PAS-ELISA against three badnavirus antisera, namely DaBV and DbBV yam badnavirus antisera (Phillips et al., 1999), as well as the general badnavirus polyclonal antisera mix 'BenL'. Spectrophotometric results showed that the antisera that cross-reacted with the maximum number of samples was 'BenL' with 90/153 (59%) positive samples, in contrast to 'DaBV' and 'DbBV' antisera which only detected 62/153 (41%) and 50/153 (33%), respectively (Fig.2, Table 1). All the samples that tested positive with DbBV, and all but two samples that tested positive by DaBV, were detected

by the BenL antisera. BenL failed to detect virus in two samples (*D. alata* Vu579a, and *D. rotundata* Benin 19) that reacted with DaBV antisera (Supplementary materials).

3.3 Immunosorbent electron microscopy (ISEM) confirmation of badnavirus particles

ISEM was performed to confirm presence of badnavirus particles in some samples. Sixteen samples were selected from the CIRAD-IRD yam collections, based on the range of PCR versus PAS-ELISA results obtained, as well as the need for >10 g leaf material to be available of samples to be used for comparative tests. Table 2 shows the direct comparison of these 16 samples by PCR, PAS-ELISA and ISEM. ISEM confirmed the presence of badnavirus particles in eight PCR- and PAS-ELISA-positive samples (*D. alata* Guinea164, *D. sansibarensis* B39, and *D. rotundata* Guinea35, Guinea43, Guinea44, Guinea50, Guinea84, and Guinea155). In addition, virus particles were observed in a further three samples (*D. alata* Cuba1, *D. rotundata* Guinea158 and Guinea163) which had tested negative by PAS-ELISA but positive by PCR. Low virus concentrations could be a plausible reason for the negative reactions by PAS-ELISA. The superior detection sensitivity of the PCR assay and use of concentrated preparations for detection in ISEM may have contributed to virus having been detected by these two methods in these samples. No badnavirus particles were detected in three samples (Pilimpikou samples A103, A105 and BF54), by ISEM or PAS-ELISA, yet these samples were PCR-positive. Two *D. alata* samples (CFPR1 and CFPC3) tested negative in all three tests.

Sequencing of the PCR amplicons from the Pilimpikou samples (sequence accession numbers AM503362, AM503363, AM503365, and AM503393, Bousalem et al., 2009) confirmed that they represented sequences that were typical of members of the genus *Badnavirus* (Fig. 4). Interestingly the sequences from the ISEM- and PAS-ELISA-negative Pilimpikou samples fell into two of the Bousalem et al. (2009) sequence groups (DBV-A and DBV-B), suggesting multiple eDBVs may be present in these samples.

3.4 Southern blot hybridisation to reveal nature of badnavirus sequences

The existence of PCR-positive, but ISEM/ELISA-negative results suggested that these Pilimpikou yam plants either had very low level virus titres only detected by the PCR technique, or that they contained eDBVs in their genomes. The Southern blotting technique was used with badnavirus probes to differentiate badnavirus sequences that hybridise to viral bands (~7-8 kb bands in the undigested sample, and several smaller fragments in digested samples), from the potential presence of hybridising bands >8 kb in digested yam genomic DNA samples, as the size of the latter means that they must represent eDBVs. The probes used were from *D. rotundata* sample Guinea155 (sequence Gn155Dr, accession number AM503383) and *D. sansibarensis* sample B39

(accession number DQ822073). These sequences fall into Bousalem et al., (2009) sequence groups DBV-A(B) and DBV-D (Fig. 4).

The autoradiographic images generated using probes Gn155Dr and B39 confirmed these RT-RNase H sequences to be of viral particle origin, with ~7-8 kb viral genome-sized bands hybridising to probe Gn155Dr in *D. rotundata* samples Guinea50 (Fig. 5, lanes 9 and 10), Guinea84 (Fig. 5, lanes 14 and 15) and Guinea155 (Fig. 5, lane 1 and 4), and to the B39 probe only in the *D. sansibarensis* plant 'B39'. For probe B39, no other hybridisation signals were visible to plant genomic DNA or to any band in the *D. rotundata, D. alata* or *D. nummularia* total DNAs tested (data not shown). In contrast, Southern hybridisation with probe Gn155Dr revealed high molecular weight (> 8 kb) plant genomic bands as well as viral particle bands. The plant genomic bands hybridising were detected in undigested (Fig. 5, lanes 1, 9 and 14) as well as *Sph*I-digested Guinea50, Guinea84 and Guinea155 DNAs (Fig. 5, lanes 4, 10 and 15). Similar-sized high molecular weight bands also hybridised in *Sph*I-digested *D. rotundata* samples Benin1, Benin15, and Guinea163, and in ISEM-ve (badnavirus-free) Pilimpikou yam samples A105 and BF54. These hybridisations signals to plant genome eDBV sequences are much stronger in the Pilimpikou (Fig. 5, lanes 2, 3, 5 and 6) and Benin '*D. rotundata*' samples (Fig. 5, lanes 17, 18) than in the Guinean samples (Fig. 5, lanes 4, 10, 15 and 20). Observation of lower molecular weight bands hybridising in the Pilimpikou sample DNAs digested with both *Sph*I and *Eco*RV (Fig. 5, lanes 7 and 8) supports the hybridisation conditions having been specific. Furthermore, the existence of stringent hybridisation conditions is supported by the absence of hybridising bands to probe Gn155Dr in *D. alata* samples (Benin33, CFPC3 and Guinea164) or *D. nummularia* sample Vu666Dn (Fig. 5, lanes 12, 13 and 19).

4. Discussion

4.1 Identification of endogenous Dioscorea bacilliform virus sequences (eDBV) in D. cayenensis-rotundata genomes

The data presented are the first results demonstrating the integration of badnavirus sequences in yam genomes of the *D. cayenensis-rotundata* species complex. EPRV sequences have not previously been reported for the plant family *Dioscoreaceae*, and support the proposition that EPRVs are widely distributed in plant genomes across a diverse range of plant families, including other cultivated hosts such as banana, petunia, rice, and tobacco (Geering et al., 2010; Iskra-Caruana et al., 2010). However, experiments were not performed to determine if genome integrants are complete

or partial genome segments, nor whether they are capable of *de novo* activation and any biological consequences.

Only a few EPRVs have been demonstrated to be capable of *de novo* generation of infectious particles (Chabannes et al., 2013; Gayral et al., 2008; Lockhart et al., 2000; Richert-Pöggeler et al., 2003). Indeed, the vast majority of EPRV integrants identified to date have been considered to be non-infectious (Chabannes and Iskra-Caruana, 2013; Geering et al., 2010; Mette et al., 2002; Staginnus et al., 2007) as a result of pseudogenisation; inactivating mutations accumulate after integration events, and hence the older an integration event, the more likely it is that it has become non-infectious (Geering et al., 2010; Iskra-Caruana et al, this volume). Such 'inert' EPRVs can, nevertheless, be important, as they have been proposed to potentially benefit their host by providing resistance to related viruses (Hull *et al*., 2000; Iskra-Caruana et al., 2010, this volume; Matzke et al., 2004; Mette *et al.*, 2002; Noreen *et al.,* 2007). An unusual situation therefore arises where the identified eDBV could be of either harm or benefit to *D. cayenensis-rotundata* breeding lines through either these integrated sequences being infectious, or by protecting their yam host from infection by closely related badnaviruses.

The eDBV confirmed here by DNA hybridisation was present in a selection of the West African samples, namely in *D. rotundata* from Guinea (Guinea-33, -50, -84, -155, and -163), Benin (Benin-1 and -15) and Burkina Faso (A103, A105 and BF54) (Fig. 5). It was not detected in the *D. alata* and *D. nummularia* samples included in Southern hybridisation studies. The probe sequence (Gn155Dr) that hybridised to this eDBV falls into species group DBV-A(B) (Bousalem et al 2009), which is a neighbouring subclade to DBV-A(A) that contains yam badnavirus DBALV for which particles have been isolated (Briddon et al., 1999; Phillips et al., 1999). In our study, badnaviral particles of DBV-A(B) occurred in addition to the plant genome eDBV sequence in the Guinean *D. rotundata* samples tested (Guinea-33, -50, -84, and -155). In contrast, the Benin *D. rotundata* and Burkina Faso Pilimpikou yam samples only showed hybridising eDBV (plant genomic) bands, with no hybridising bands in the viral particle range. This demonstrates that sequences falling into putative species group DBV-A (B) are integrated in the plant genome, but also occur in episomal virus genomes.

Southern hybridisation studies on DBSNV showed probe 'B39' to hybridise only to viral genome originating from virus particles (~7-8 kb) in total DNAs of the *D. sansibarensis* sample, and did not hybridize with host genome sequences of the West African *D. rotundata* or *D. alata* samples. This suggests there are no 'DBSNV-like' eDBV sequences in the *D. sansibarensis*, *D. rotundata* or *D. alata* genomes analyzed here. A similar observation has been made in banana, in which BSV sequences originating from only viral particles have been demonstrated to form a phylogenetically distinct

clade (Clade 3), separated from endogenous badnavirus sequences with similarities to episomal BSV genomes (Clade 1) or without corresponding episomal forms (Clade 2) (see Iskra-Caruana et al., this volume).

4.2 Is the identified eDBV infectious?

Initially it was postulated that the 100% PCR-positive results for the *D. cayenensis-rotundata* samples in the CIRAD-Guinea collection, combined with 94% PAS-ELISA positives (Fig. 2), indicated that an activatable EPRV was leading to infection by a single badnavirus in all samples from the Republic of Guinea, mirroring the situation that led to the first discovery of an endogenous and activatable ('Clade 1') eBSV in *Musa* sp. (Harper et al., 1999a). However, the variable reactions of the 77 Guinea samples to the three badnavirus antisera (Table 2, and Supplementary materials) did not support this, as they suggested not all samples were infected with the same virus. Sequence data confirmed this with Guinean samples showing considerable nucleotide variation (identity levels as low as 64.1%) and falling into three of the Bousalem et al., (2009) putative species groups, namely DBV-A, DBV-B and DBV-D (Fig. 4). Southern hybridisation data confirmed that diverse viruses were present and that presence of eDBV was not always associated with the presence of viral particles that would fall into the DBV-A(B) species group. Viral and plant genome bands hybridised to Gn155Dr in some samples (Guinea-33, -50, -84 and -155), but only plant genome bands hybridised in Guinea163, Benin1 and Benin15 (Fig. 5), despite ISEM and PAS-ELISA having revealed that the latter three samples also contained badnaviral particles (Table 2, Supplementary Materials).

The inability to detect any viral particles in Pilimpikou samples using a range of techniques, highlights that the eDBV(s) identified have not generated episomal viral infections in this genetic background. This is unsurprising, as EPRVs in general are unable to give rise to functional genomes (Kunii et al., 2004; Geering et al. 2005). In fact infectious EPRVs are the exception, as illustrated by banana, in which to date 27 separate BSV integration events have been discovered of which only three represent infectious EPRVs (eBSVs; Gayral et al., 2010; Gayral and Iskra-Caruana 2009).

Further studies are needed to determine if the viral particle sequences hybridising to probe Gn155Dr in Guinea *D. rotundata* samples (Guinea50, Guinea84 and Guinea155) have arisen from the eDBV sequences in these samples. It is not clear currently whether this identified eDBV is infectious and hence poses a threat to yam breeding programmes. Further characterisation studies on this eDBV are being performed to establish its relatedness to viral particles present in *D. cayenensis-rotundata* material (Marie Umber et al., unpublished data). Such studies are also investigating the

'DBV-B' RT-RNaseH sequences in the Pilimpikou samples, (Fig. 4) which appear to represent another eDBV integration event.

If eDBV(s) are shown to be activatable, it will be critical to determine accurately the ploidy levels and phyletic relationships of the yam genomes of the samples from Guinea, Benin and Burkina Faso to improve our understanding of which genetic backgrounds eDBV(s) can lead to episomal viral infections and to what extent the findings can be related to knowledge on eBSVs (Chabannes et al., 2013; Iskra-Caruana et al., 2010; 2013).

4.3 Implications of eDBV for interpretation of existing yam badnavirus diagnostic results

Diagnostic techniques are essential to detect badnavirus infections in yams, particularly as infections are often present without any marked symptoms (Fig. 1b) (Kenyon et al., 2008; Seal and Muller, 2007). The serological heterogeneity of badnaviruses, combined with their tendency to develop low titre, has meant that PCR amplification has been a favoured method for their detection in hosts lacking EPRVs (Muller et al., 2011; Yang et al., 2003). However, the identification of eDBVs in the *D. cayenensis-rotundata* material examined here, combined with the 100% PCR-positive results obtained for the 156 *D. cayenensis-rotundata* samples (29 CIRAD-IRD collection, 112 IITA breeding lines, and 15 IITA landraces) tested, highlights that PCR is of little practical use for virus-indexing yams belonging to this species complex.

Of the three antisera used here, only the BenL antisera appears of value for estimating the prevalence of badnaviruses. This antisera detected badnavirus infections in 71 of 77 samples (92%) from Guinea, in contrast to only 47 (61%) and 44 (57%) using the DaBV and DbBV antisera respectively. Previous reports of 45% badnavirus infection in yam field surveys in Benin (Eni et al., 2008), and 35% badnavirus infection in ~700 South Pacific yam samples (Kenyon et al., 2008) are likely to be underestimates, as both these studies used the same DaBV polyclonal antiserum.

The BenL antisera is a mix of ~30 different polyclonals to sugarcane bacilliform viruses and BSVs (Le Provost et al., 2006). Interestingly, although it was not designed to target yam badnaviruses, it was a more effective antiserum for detecting yam badnavirus infections than the two yam badnavirus polyclonals DaBV and DbBV designed to two yam badnavirus types with differing properties (Phillips et al., 1999). The BenL antisera, nevertheless, missed two infections (Vanuatu *D. alata* sample Vu579a, and *D. rotundata* sample Benin19) detected using the DaBV antiserum. Furthermore there were an additional 26 of the 153 samples screened by PAS-ELISA that were 'BenL-ve', yet PCR-positive (Supplementary materials). Six of these were examined by ISEM, and it was shown that even the best antisera BenL had failed to detect badnavirus particles that

occurred in three of these samples (Table 2). None of the three antisera screened can therefore provide a confident measure of virus status to use for indexing yam plants for badnavirus infection.

The three 'PCR-positive but PAS-ELISA-ve' samples in which badnavirus particles were detected by ISEM originated from *D. rotundata* (Guinea158, Guinea163) and *D. alata* (Cuba1) (Table 2). Phylogenetic analysis of previously determined nucleotide sequence data from these samples clusters sequences into the putative sequence groups DBV-A(A) (Guinea163), DBV-B (Cuba1 and Guinea158), DBV-D (Guinea163) (Fig.4). It seems probable that the DBV-B and DBV-D sequence groups present two additional virus species of the genus *Badnavirus* where episomal virus exists, in addition to the ICTV-recognised species DBALV and DBSNV (King et al., 2012). It appears these species do not cross-react with any of the current antisera, and this is a possible cause for the large discrepancy in PCR (50%) and PAS-ELISA (9%) results for *D. alata* samples from C./S. America.

The identification of eDBVs in yam, combined with the lack of antisera that cross-react sufficiently with all virus strains by ELISA, meant that ISEM using partially purified preparations was essential in this study to determine whether viral particles occurred in PCR-positive, but ELISA-negative samples. However, ISEM is not suited for routine virus-indexing, as it is a highly time-consuming technique that for low titre badnavirus infections requires semi-purified virus preparations. The latter requires much greater quantities of leaf material, which may not be available, particularly if the material to be screened is from tissue culture. The most labour-intensive part of the process is the thorough visual examination of grids required to ensure infections are not missed, particularly if the virus does not cross-react well to the antisera used, as observed for *D. alata* sample Cuba1 in this study.

4.4 Diagnostics to distinguish eDBVs from DBVs

Rapid diagnostic techniques are required that will allow the differentiation of *D. cayenensis-rotundata* material containing only non-functional eDBVs from those in which viral particles occur. Although yam has to date been an under-researched crop, the similar challenges faced for indexing banana (*Musa* sp.) for BSVs, means that techniques developed in this field should be transferable to yam.

Immunocapture (IC)-PCR was the first technique described as ideal for distinguishing BSV particles from EPRVs (Harper et al., 1999b). However, for yam using the BenL antisera, IC-PCR gave non-reproducible results which were considered to be due to yam extracts containing high levels of polysaccharides, as well as contamination with plant genomic DNA (S. Seal, unpublished data). Le Provost et al. (2006) reported these limitations also to be experienced when using either IC-PCR or

direct binding PCR for episomal BSV detection. They devised an improved protocol to avoid spurious positive results arising from contaminating plant genomic DNA binding to PCR tubes. This improved method holds promise for rapid testing of yam samples, but will not be of diagnostic value until a more specific antiserum has been developed that cross reacts to all yam badnaviruses. Le Provost et al. (2006) also highlighted that species-specific primers were needed for episomal BSV detection, as the degenerate primer sets used on DNA extracts were not sufficiently sensitive in an immunocapture-PCR format due to low virus titres and the high genetic variability in BSV.

Real-time PCR methods have been described able to distinguish eBSV sequences from those of episomal BSV (Delanoy et al., 2003), and with real-time PCR now becoming more routine in Africa, this technique offers promise to distinguish eDBVs from sequences originating from viral particles. A further method reported as being able to distinguish viral particles from EPRVs is rolling circle amplification (RCA) (Rector et al., 2004), but our experience with yam indicates that it cannot be used for rapid indexing purposes as it amplifies plant sequences at low frequency. The size of these can be confused with the bands generated for viral particles (unpublished data). Although not suitable as a rapid virus-indexing method, the RCA technique is useful for research purposes and to demonstrate the episomal status of EPRVs, as performed in banana by James et al., (2011).

The introduction of improved diagnostics to distinguish DBVs from eDBVs therefore depends on the development of improved antisera, as well as further characterisation of eDBV sequences and such work is in progress at CIRAD, IITA and NRI. This work will be aided considerable by a draft yam (*D. rotundata*) genome which is expected to be available shortly (Tamiru et al., 2013).

5. Conclusions

This first discovery of integrated badnavirus sequences (eDBVs) in yam (*Dioscorea* spp.) has led to the realisation that PCR diagnostic techniques are not adequate for enabling decisions to be made on the suitability of yam germplasm for wide-scale cultivation or international exchange between yam breeding programmes. This study has also demonstrated that existing serological techniques for yam badnaviruses are inadequate in failing to cross-react sufficiently to some isolates. These findings are significant in light of the extremely high prevalence of badnavirus infections noted in West Africa (e.g. >94% in Guinean *D. cayenensis-rotundata* samples in this study), and an appreciation that lack of virus-free planting material is a major constraint to improvement of yam yields in this region (IITA, 2012).

Further research is needed to characterise the eDBVs identified and determine their prevalence. The 100% PCR-positive results obtained for the 156 *D. cayenensis-rotundata* samples

tested in this study, together with sequence and hybridisation studies, suggest that they may exist in all germplasm that falls into this yam species complex. Studies on whether the eDBVs are infectious will be essential to determine their biological and epidemiological significance; do these sequences act as sources of infection, and/or do they provide resistance to homologous badnavirus through gene silencing mechanisms (Iskra-Caruana et al., 2010, this volume; Mette et al. 2002)? For these studies, improved diagnostic tools are needed, such that antisera detect all badnavirus particles and PCR primers distinguish sequences present in these particles from eDBVs.

Until the development of improved diagnostic tools for yam badnaviruses, it is suggested that the approach used in our study will be valuable for breeding programmes. Use of ELISA with existing antisera as well as PCR will allow the confirmation of some PCR results. PCR-positive but ELISA-negative results can then be studied in more detail to determine whether viral particles are present. The rapidly decreasing costs of next generation sequencing (NGS) may mean that this will become the most reliable method of screening promising yam breeding lines or landraces for their viral and eDBV content prior to 'seed yam' multiplication. This technology has already been validated as useful for the characterisation of unknown viruses (Adams et al., 2009), and will become particularly useful when a draft yam genome is available to simplify bioinformatic analyses of NGS data.

Once yam lines are identified that are free of virus and infectious eDBVs, the next challenge will be to develop a strategy for their distribution to yam farmers. The current extremely high prevalence of badnavirus infections in some regions highlight that for the introduction of 'clean' seed to have maximum impact in reducing virus incidence, careful management will be required to prevent virus-free improved material becoming infected rapidly when planted in the field. The challenges for research aimed at improving yam yields, and hence food security in West Africa therefore remain great.

Acknowledgement

The authors thank Mustapha Bousalem and Camara Fadjimba for providing access to yam plants collected and maintained at CIRAD, Montpellier. Funding for this project was from a combination of sources; authors were generally 'core'-funded by their associated employing institutions, with the exception of Aliyu Turaki who received funding from Kebbi State Government, Nigeria as well as partial research support funding from a Bill and Melinda Gates Foundation-funded project entitled 'Yam Improvement for Incomes and Food Security in West Africa (YIIFSWA, 2011-2016) granted to IITA, Nigeria.

References

Adams, I.P., Glover, R.H., Monger, W.A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M., Boonham, N., 2009. Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. Molecular Plant Pathology 10, 537-545.

Asiedu, R., Sartie, A., 2010. Crops that feed the World, 1. Yams. Food Security 2, 305-315.

Bousalem, M., Durand, O., Scarcelli, N., Lebas, B.S.M., Kenyon, L., Marchand, J.-L., Lefort, F., Seal, S.E., 2009. Dilemmas caused by endogenous pararetroviruses regarding the taxonomy and diagnosis of yam (*Dioscorea* spp.) badnaviruses: analyses to support safe germplasm movement. Archives of Virology 154, 297-314.

Briddon, R.W., Phillips, S., Brunt, A., Hull, R., 1999. Analysis of the sequence of Dioscorea alata bacilliform virus; comparison to other members of the badnavirus group. Virus Genes 18, 277-283.

Chabannes, M., Baurens, F.C., Duroy, P.-O., Sidibe-Bocs, S., Vernerey, M.-S., Rodier-Goud, M., Barbe V., Gayral, P., Iskra-Caruana, M.-L., 2013. Three infectious viral species lying in wait in the banana genome. Journal of Virology 87, 8624-8637.

Chabannes, M., Iskra-Caruana, M.-L., 2013. Endogenous pararetroviruses-a reservoir of virus infection in plants. Current Opinion in Virology. doi:pii: S1879-6257(13)00148-X. 10.1016/j.coviro.2013.08.012

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

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

Delanoy, M., Salmon, M., Kummert, J. 2003. Development of real-time PCR for the rapid detection of episomal *Banana streak virus* (BSV). Plant Disease 87, 33-38.

Dumont, R., Dansi, A., Vernier, P., Zoundjihekpon, J., 2006. Biodiversity and domestication of yams in West Africa. Traditional practices giving rise to *Dioscorea rotundata* Poir. Editions Quae, CIRAD-IPGRI, Nancy, France.

Eni, A.O., Hughes, J.d'A., Asiedu, R., Rey, M.E.C., 2008a. Sequence diversity among badnavirus isolates infecting yam (*Dioscorea* spp.) in Ghana, Togo, Benin and Nigeria. Archives of Virology 153, 2263-2272.

Eni, A.O., Hughes, J.d'A., Rey, M.E.C., 2008b. Survey of the incidence and distribution of five viruses infecting yams in the major yam-producing zones in Benin. Annals of Applied Biology 156, 223-232.

Eni, A.O., Hughes, J.d'A., Asiedu, R., Rey, M.E.C., 2009. Survey of the incidence and distribution of viruses infecting yam (*Dioscorea* spp.) in Ghana and Togo. Annals of Applied Biology 156, 243-251.

F.A.O., 2012. Food and Agricultural Organisation of the United Nations FAO Statistics 2012. Rome, Italy. [available online] at www.faostat.fao.org.

Galzi S., Ricot S., Emmanuel P., Philippe R., Filloux, D., 2013. Assessment and characterisation of the diversity of viruses infecting cultivated yams (*Dioscorea* spp.) in Haïti. Abstract in the 14íémes Rencontres de Virologie Végétale meeting, 13 - 17 January 2013 held in Aussois, France.

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

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

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

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

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

Goudou-Urbino, C., Konate, G., Quiot, J.B., Dubern, J., 1996. Aetiology and ecology of a yam mosaic disease in Burkina Faso. Tropical Science 36, 34-40.

Harper, G., Ganesh, D., Thottappilly, G., Hull, R., 1999b. Detection of episomal Banana streak virus by IC-PCR. Journal of Virological Methods 79, 1–8.

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

Harper, G., Hull, R., Lockhart, B.E.L, Olszewski, N.E., 2002. Viral sequences integrated into plant genomes. Annual Review of Phytopathology 140, 119-136.

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

Harrison, B.D., Roberts, I.M., 1973. Association of virus-like particles with internal brown spot of yam (*Dioscorea alata*). Tropical Agriculture, University of West Indies, IPC, volume 50, 335-340.

IITA, 2012. New project to invest US$12 million to boost yam productivity in Ghana and Nigeria 02/04/2012 Press release-IITA [available online] at www.iita.org/news-frontpage-feature2.

Iskra-Caruana, M.-L., Baurens, F.C., Gayral, P., Chabannes, M., 2010. A four-partner plant-virus interaction: enemies can also come from within. Molecular Plant Microbe Interactions 23, 1394-1402.

Iskra-Caruana, M.-L., Chabannes, M., Duroy, P.O., Muller, E., this issue. A possible scenario for the evolution of *Banana streak virus* in banana. Virus Research

Iskra-Caruana, M.-L., Duroy, P.-O., Chabannes, M., Muller, E., 2013. The common evolutionary history of badnaviruses and banana. Inf. Genetics Evol. in press

Jakowitsch, J., Mette, M.F., van der Winden, J., Matzke, M.A., Matzke, A.J.M., 1999. Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. Proceedings of the National Academy of Sciences USA 96, 13241-13246.

James, A.P., Geijskes, R.J., Dale, J.L., Harding, R.M., 2011. Development of a novel rolling-circle amplification technique to detect Banana streak virus that also discriminates between integrated and episomal virus sequences. Plant Disease 95, 57-62.

Kenyon, L., Lebas, B.S.M., Seal, S.E., 2008. Yams (*Dioscorea* spp.) from the South Pacific Islands contain many novel badnaviruses: implications for international movement of yam germplasm. Archives of Virology 153, 877-889.

Kenyon, L., Shoyinka, S.A., Hughes, J.d'A., Odu, B.O., 2001. An overview of viruses infecting *Dioscorea* yams in Sub-Saharan Africa. In: Plant Virology in Sub-Saharan Africa. Eds. Hughes Jd'A, Odu BO. pp. 432-439.

King, A.M.Q., Adams, M.J., Carstens, E.B., , Lefkowitz, E.J., 2012. Virus taxonomy, Ninth report of the International Committee on Taxonomy of Viruses, San Diego, Elsevier. 1338 pp.

Kunii, M., Kanda, M., Nagrano, H., Uyeda, I., Kishima Y., Sano, Y., 2004. Reconstruction of putative DNA virus from endogenous rice tungro bacilliform virus-like sequences in the rice genome: implications for integration and evolution. BMC Genomics 5, 80.

Le Provost, G., Iskra-Caruana, M.-L., Acina, I., Teycheney, P.-Y. 2006. Improved detection of episomal banana streak viruses by multiplex immunocapture PCR. Journal of Virological Methods 137, 7-13.

Lockhart, B. E. L., 1986. Purification and serology of a bacilliform virus associated with banana streak disease. Phytopathology 76, 995-999.

Lockhart, B.E.L., Menke, J., Dahal, G., Olszewski, N.E., 2000. Characterization and genomic analysis of tobacco vein clearing virus, a plant pararetrovirus that is transmitted vertically and related to sequences integrated in the host genome. Journal of General Virology 81, 1579-1585.

Lodhi, M.A. Guang-Ning, Y., Norman, F.W., Bruce, I.R., 1994. A simple and efficient method for DNA extraction from grapevine cultivars, Vitis species and Ampelopsis. Plant Molecular Biology Reporter 12, 6-13.

Mantell, S.H., Haque, S.Q., 1978. Incidence of internal brown spot disease in white Lisbon Yams (*Dioscorea alata*) during storage. Experimental Agriculture 14, 167-172.

Matzke, M., Gregor, W., Mette, M.F., Aufsatz, W., Kanno, T., Jakowitsch, J., Matzke, A.J.M., 2004. Endogenous pararetroviruses of allotetraploid *Nicotiana tabacum* and its diploid progenitors, *N. sylvestris* and *N. tomentosiformis*. Biological Journal of the Linnean Society, 82: 627–638

Mette, M.F., Kanno, T., Aufsatz, W., Jakowitsch, J., van der Winden, J., Matzke, M.A., Matzke, A.J.M., 2002. Endogenous viral sequences and their potential contribution to heritable virus resistance in plants. EMBO Journal 21, 461-469.

Mignouna, H.D., Abang M.M., Asiedu, R., 2008. Genomics of yams, a common source of food and medicine in the tropics pp. 549-570. In Plant Genetics and Genomics: Crops and Models. Moore P. and Ming R. (eds.), Springer Berlin, Germany.

Muller, E., Dupuy, V., Blondin, L., Bauffe, F., Daugrois, J.-H., Laboureau, N., Iskra-Caruana, M.-L., 2011. High molecular variability of sugarcane bacilliform viruses in Guadeloupe implying the existence of at least three new species. Virus Research 160, 414-419.

Mumford, R.A., Seal, S.E., 1997. Rapid single-tube immunocapture RT-PCR for the detection of two yam potyviruses. Journal of Virological Methods 69, 73-79.

Ndowora, T., Dahal, G., LaFleur, D., Harper, G., Hull, R., Olszewski, N.E., Lockhart B.E.L., 1999. Evidence that badnavirus infection in Musa can originate from integrated pararetroviral sequences. Virology 255, 214–220.

Perrier, X., Flori, A., Bonnot, F., 2003. Data analysis methods. In: Hamon, P., Seguin, M., Perrier, X. Glaszmann, J.C. (eds) Genetic diversity of cultivated tropical plants. Enfield Science Publishers, Montpellier, pp 43-76.

Phillips, S., Briddon, R.W., Brunt, A.A., Hull, R., 1999. The partial characterization of badnavirus infecting the Greater asiatic or water yam (*Dioscorea alata*). Journal of Phytopathology 147, 265-269.

Rector, A., Tachezy, R., Van Ranst, M., 2004. A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. Journal of Virology 78, 4993-4998.

Richert-Pöggeler, K.R., Noreen, F., Schwarzacher, T., Harper, G., Hohn, T., 2003. Induction of infectious petunia vein clearing (pararetro)virus from endogenous provirus in petunia. EMBO Journal 22, 4836-4845.

Scarcelli N., Tostain S., Mariac C., Agbangla C., Ogoubi D., Julien B., Pharm J-L., 2006. Genetic nature of yams (*Dioscorea* sp.) domesticated by farmers in Benin (West Africa). Genetic Resources and Crop Evolution 53, 121-130.

Seal, S.E., Muller, E., 2007. Molecular analysis of a full-length sequence of a new yam badnavirus from *Dioscorea sansibarensis*. Archives of Virology 152, 819-825.

Staginnus, C., Iskra-Caruana, M., Lockhart, B., Hohn, T., Richert-Pöggeler, K.R., 2009. Suggestions for a nomenclature of endogenous pararetroviral sequences in plants. Archives of Virology 154, 1189-1193.

Staginnus, C., Gregor, W., Mette, M.F., Teo, C.H., Borroto-Fernández, E.G., Machado, M.L., Matzke, M., Schwarzacher, T., 2007. Endogenous pararetroviral sequences in tomato (*Solanum lycopersicum*) and related species. BMC Plant Biology, 7, 24

Tamiru, M., Natsume, S., Takagi, H., Babil, P., Yamanaka, S., Lopez-Montes, A., Gedil, M., Bhattacharjee, R., Takagi, H., Asiedu, R., Terauchi, R., 2013. Whole genome sequencing of Guinea yam (*Dioscorea rotundata*). Abstract, p20 In Book of abstracts for Yams-2013, First global conference on Yam, 3-6 October 2013, Accra, Ghana

Wylie, S., Wilson, C.R., Jones, R.A.C., Jones, M.G.K., 1993. A polymerase chain reaction assay for cucumber mosaic virus in lupin seeds. Australian Journal of Agricultural Research 44, 41-51.

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

Fig. 1. Variable symptoms on yam leaves infected with badnaviruses. Leaves of a) *D. rotundata* (var. Nwokpoko) showing veinal chlorosis, b), *D. rotundata* (var. Ogoja) showing no marked viral symptoms.

Fig. 2. Percentage of badnavirus ELISA- and PCR-positive results of samples tested from CIRAD-IRD yam collections maintained in Montpellier, France. ELISA results shown are for overnight readings using three different polyclonal antisera to badnaviruses (BenL, DaBV and DbBV). For PCR, numbers of samples tested were CIRAD-IRD Africa (n=11), CIRAD-IRD C/S America (n=15), CIRAD-IRD-Guinea (n=14), CIRAD-IRD Vanuatu (n=9), and CIRAD-IRD Benin (n=10). For ELISA screening, sample numbers were CIRAD-IRD Africa (n=11), CIRAD-IRD C/S America (n=15), CIRAD-IRD-Guinea (n=77), CIRAD-IRD Vanuatu (n=27), and CIRAD-IRD Benin (n=23). Details of the origin of individual samples and their individual badnavirus diagnostic PCR and ELISA test results are given in the Supplementary materials.

Fig. 3. Detection of badnavirus sequences by PCR in *Dioscorea rotundata* breeding lines (IITA) using primer set Badna-FP/-RP (Yang et al., 2003) that generates a PCR product of ~579 bp. Lane M = 100 bp marker (New England Biolabs, UK), w = water (no added template) negative control, lanes 1-14, represent IITA breeding lines, where 1= B12b (TDr 97/00917 × TDr 99/02607), 2= B13b (TDr 04/219 × TDr 98/02677), 3= B16a (TDr 97/00917 × POUNA), 4= B19a (TDr 89/02475 × TDr 97/00777), 5=B20a (TDr 99/02793 × TDr 1892), 6=B21c (TDr 04/219 × TDr 04/219), 7=B22c (TDr 04/219 × TDr 97/00777), 8= B17d (TDr 97/00205 ×TDr 1892), 9= B23e (TDr 97/00917 × TDr 99/02607), 10= B26d (TDr 99/02793 OP),11= B28d (TDr 95/18544 OP), 12= B29e (TDr 96/00629 OP), 13= B30d (TDr 97/00917 OP), 14= B31e (TDr 96/00639 OP).

Fig. 4. Neighbour-Joining phylogenetic tree showing the relationships of nucleotide sequences (in bold lettering) from the yam samples used for Southern hybridisation with two different probes Gn155Dr and B39 (sequences underlined). Included in the analysis are equivalent 529 bp RT-RNaseH sequences of DBALV (X94576), CSSV (AJ781003), BSOLV (AJ002234), ComYMV (NC001343), SCBMOV (M89923), TaBV (AF357836) and outgroup RTBV (X57924), as well as representative sequences of all monophyletic groups described by Bousalem et al., (2009) where DBV-A = Dioscorea bacilliform virus A (A and B subgroups), DBV-B = Dioscorea bacilliform virus B, DBV-C = Dioscorea bacilliform virus C, DBV-D = Dioscorea bacilliform virus D, DeBV-A = Dioscorea esculenta bacilliform virus A, DeBV-B = Dioscorea esculenta bacilliform virus B, DeBV-C = Dioscorea esculenta bacilliform virus C, DeBV-D = Dioscorea esculenta bacilliform virus D, DeBV-E = Dioscorea esculenta bacilliform virus E, DeBV-F = Dioscorea esculenta bacilliform virus F, and DpBV = Dioscorea pentaphylla bacilliform virus. The eleven corresponding Kenyon et al., (2008) groupings to these monophyletic groups are also given and denoted by K1-K11. The bootstrap analysis of the sequences was 1,000 replicates and the cut-off value was 70%.

Fig. 5. Autoradiograph of Southern blot of total uncut and restriction enzyme-digested *Dioscorea* spp. DNA samples, revealing hybridisation of *D. cayenensis-rotundata* (*Dcr*) genomic DNA to badnavirus RT-RNaseH sequence probe Gn155Dr. Following hybridisation with the ^{32}P-labelled Gn155Dr sequence, washes were performed under stringency conditions (0.2x SSC, 60^{o}C) to achieve hybridisation only to sequences of the same 'DBV' species group. Lanes represent 1= *Dcr* Guinea155 uncut, 2 = *Dcr* A105 *Sph*I-digested, 3 = *Dcr* BF54 *Sph*I-digested, 4 = *Dcr* Guinea155 *Sph*I-digested, 5 = *Dcr* A105 *Sph*I-digested, 6 = *Dcr* BF54 *Sph*I-digested, 7 = *Dcr* A105 *Sph*I+*Eco*RV-digested, 8 = *Dcr* BF54 *Sph*I+*Eco*RV-digested, 9 = *Dcr* Guinea50 uncut, 10 = Guinea50 *Sph*I-digested (incomplete digestion), 11 = Guinea50 *Sph*I+*Eco*RV-digested, 12 =*D. alata* CFPC3 *Sph*I-digested, 13 = *D. nummularia* Vu666n *Sph*I-digested, 14 = *Dcr* Guinea84 uncut, 15 = *Dcr* Guinea84 *Sph*I-digested, 16, low concentration DNA sample , 17= *Dcr* Benin1 *Sph*I-digested, 18 = *Dcr* Benin15 *Sph*I-digested, 19 = *D. alata* Benin33 *Sph*I-digested, 20 = *Dcr* Guinea163 *Sph*I-digested, 21 = *D.alata* Guinea164 *Sph*I-digested. The top two filled arrows show hybridising endogenous Dioscorea bacilliform virus (eDBV) sequences present in the plant genomes (focussing on the faint band in lane 4, 10 and 15), whereas the lower two 'open' arrows indicate the size range 7-8 kb and show uncut/closed circular virus genome bands (at ~7 kb in lanes 1, 9 and 14), and linearised/open circular virus bands (at ~8 kb in lanes 4, 10, and 15). There is undigested DNA present in lane 4 and lane 10, with lane 15 showing the complete digestion pattern. All lane DNAs were of approximately the same concentration (~10 ug per lane), except for lanes 1,9, 10 and 11 being slightly higher, and lane 16 being of very low concentration.

Table 1

Summary of badnavirus PCR- and ELISA-positive results obtained for samples within the individual CIRAD yam collections, with a breakdown according to Dioscorea species. ELISA results shown are for readings after 18h at 37°C using three different polyclonal antisera to badnaviruses (BenL, DaBV and DbBV). Details of the origin of individual samples and their individual badnavirus diagnostic PCR and ELISA test results are given in the Supplementary materials. A few samples were not tested (nt) due to poor DNA extractions.

CIRAD Yam collection	**Badna PCR**	**BenL**	**DaBV**	**DbBV**
CIRAD IRD Africa Total consisting of:	***11/11***	***4/11***	***2/11***	***2/11***
D. cayenensis-rotundata (incl. Pilimpikou)	7/7	1/7	1/7	1/7
D. dumetorum	2/2	2/2	0/2	0/2
D. esculenta	1/1	0/1	0/1	0/1
D. sansibarensis	1/1	1/1	1/1	1/1
CIRAD IRD C./S. America Total consisting of:	***9/14***	***1/15***	***0/15***	***0/15***
D. alata	5/10	1/11	0/11	0/11
D. trifida	4/4	0/4	0/4	0/4
CIRAD-Guinea Total consisting of	***14/14***	***71/77***	***47/77***	***44/77***
D. cayenensis-rotundata	13/13	68/72	45/72	42/72
D. alata	1/1	1/1	1/1	1/1
D. abyssinica	nt	2/4	1/4	1/4
CIRAD-Vanuatu Total consisting of	***3/9***	***5/27***	***3/27***	***2/27***
D.alata	1/4	5/22	3/22	2/22
D.nummularia	2/4	0/5	0/5	0/5
CIRAD-Benin consisting of	***10/10***	***9/23***	***10/23***	***2/23***
D.alata	1/1	0/1	0/1	0/1
D.cayenensis-rotundata	9/9	9/22	10/22	2/22
Total	**47/58**	**90/153**	**62/153**	**50/153**
	(81%)	**(59%)**	**(41%)**	**(33%)**

Table 2

Comparison of PCR diagnostic results for yam samples with their reaction through PAS-ELISA to three polyclonal antisera (BenL, DaBV, and DbBV) and the presence of particles detected by immunosorbent electron microscopy (ISEM, using the BenL polyclonal mix as the coating antisera).

Sample	Origin	*Dioscorea* sp.	PCR	Sequence groups*	ISEM	BenL ELISA	DaBV ELISA	DbBV ELISA
A103	Burkina Faso	*D. cayenensis-rotundata* Pilimpikou	+	DBV-A(A), DBV-B	-	-	-	-
A105	Burkina Faso	*D. cayenensis-rotundata* Pilimpikou	+	DBV-B	-	-	-	-
BF54	Burkina Faso	*D. cayenensis-rotundata* Pilimpikou	+	n/a	-	-	-	-
CFPR1	Martinique	*D. alata*	-	n/a	-	-	-	-
CFPC3	Martinique	*D. alata*	-	n/a	-	-	-	-
B39	Benin	*D. sansibarensis*	+	DsBV, DBV-B	++	++	++	+++
Cuba1	Cuba	*D. alata*	+	DBV-B	+	-	-	-
Guinea35	Guinea	*D. rotundata*	+	n/a	+	+++	+++	+++
Guinea43	Guinea	*D. rotundata*	+	DBV-A	+	++	+	+
Guinea44	Guinea	*D. rotundata*	+	DBV-A	+	++	+	+
Guinea50	Guinea	*D. rotundata*	+	DBV-A(A), DBV-A(B)	++	++	++	++
Guinea84	Guinea	*D. rotundata*	+	DBV-A(A), DBV-A(B)	++	+	++	++
Guinea155	Guinea	*D. rotundata*	+	DBV-A(B)	++	+	-	-
Guinea158	Guinea	*D. rotundata*	+	DBV-B	++	-	-	-
Guinea163	Guinea	*D. rotundata*	+	DBV-A(A), DBV-D	++	-	-	-
Guinea164	Guinea	*D. alata*	+	DBV-D	++	+	+	+

* = clones from PCR products in some instances represented more than one of the sequence groups described by Bousalem et al., (2009)

n/a = not applicable, as no PCR product generated, or failure to clone PCR product