Characterization and genetic diversity of Dioscorea bacilliform viruses present in a Pacific yam germplasm collection

Amit C. Sukal1,2 | Dawit B. Kidanemariam1 | James L. Dale1 | Rob M. Harding1 | Anthony P. James1

Abstract
Yam is an important food staple in many Pacific Island countries, but its production is limited by several factors including virus infection. The presence of viruses also hinders the exploitation of the crop by restricting the international dissemination of yam germplasm due to quarantine concerns. In this study, a total of 224 yam accessions present in the Pacific germplasm collection at SPC-CePaCT, Fiji, were tested for the presence of badnaviruses using a combination of rolling circle amplification (RCA) and sequencing. Of these, 35 tested positive for either Dioscorea bacilliform AL virus (DBALV) or Dioscorea bacilliform AL virus 2 (DBALV2). DBALV was only detected in samples from Vanuatu and Tonga, while DBALV2 was restricted to Papua New Guinean samples. Ten complete genome sequences of DBALV and DBALV2 were obtained and pairwise comparison of the partial reverse transcriptase/ribonuclease H (RT/RNase H)-coding sequence revealed identities ranging from 89% to 100% and 92% to 100%, respectively. The results suggest that DBALV and DBALV2 are the prevalent badnavirus species infecting yams in the Pacific region.

KEYWORDS
badnavirus, DBALV, DBALV2, Dioscorea bacilliform virus, next generation sequencing, yam

1 | INTRODUCTION

Yam (Dioscorea spp.) provides a staple food source for millions of people in Africa, South America, Asia, and the Pacific as well as having significant cultural and economic importance (Bourke and Vlassak, 2004; Sukal et al., 2015). While Africa accounts for 95% of total world production, predominated by Dioscorea rotundata-cayenensis (FAOSTAT, 2018), yam is also very important in the Pacific, where D. alata and D. esculenta are the dominant species (Kenyon et al., 2008). Yam is almost exclusively vegetatively propagated, which facilitates the vertical transmission of viruses leading to virus accumulation and associated production losses. Viruses belonging to the families Alphaflexiviridae (genus Potexvirus), Betaflexiviridae (genus Carlaviridae), Bromoviridae (genus Cucumovirus), Caulimoviridae (genus Badnavirus), Potyviridae (genus Macluravirus and Potyvirus), Secoviridae (genus Comovirus and Fabavirus) and Tombusviridae (genus Aureusvirus) are known to infect yams (Bömer et al., 2019; Silva et al., 2019). As such, the presence of viruses presents major challenges for the safe exchange of yam germplasm (Kenyon et al., 2008; Seal et al., 2014; Sukal et al., 2015).

Badnaviruses are the most widespread viruses known to infect yams and are classified as a number of Dioscorea bacilliform virus (DBV) species (Eni et al., 2008; Kenyon et al., 2008; Bousalem et
DBVs are transmitted vegetatively as well as by various mealybugs (family Pseudococcidae) in a semipersistant manner (Phillips et al., 1999; Odu et al., 2004). Although infection of yams with DBVs has been associated with leaf symptoms such as veinal chlorosis, necrosis, puckering, and crinkling, symptomless infections can also occur (Phillips et al., 1999; Lebot, 2009; Seal et al., 2014; Bömer et al., 2016).

At present, seven distinct DBV species have been taxonomically accepted by the ICTV, namely Dioscorea bacilliform AL virus (DBALV), Dioscorea bacilliform Al virus 2 (DBALV2), Dioscorea bacilliform ES virus (DBESV), Dioscorea bacilliform RT virus 1 (DBVRTV1), Dioscorea bacilliform RT virus 2 (DBVRTV2), Dioscorea bacilliform SN virus (DBSNV), and Dioscorea bacilliform TR virus (Adams and Carstens, 2012; Sukal et al., 2017; Adams et al., 2018). An additional putative species, namely Dioscorea bacilliform RT virus 3 (DBVRTV3) has also been described (Bömer et al., 2018), with additional PCRM-based studies suggesting the existence of at least another eight putative yam-associated badnavirus species worldwide (Kenyon et al., 2008; Bousalem et al., 2009; Bömer et al., 2016).

The Centre for Pacific Crops and Trees (CePaCT) within the Pacific Community (SPC) maintains a unique collection of in vitro yam germplasm from the Pacific region. Although this germplasm represents a potentially valuable resource to improve yam production, a lack of reliable diagnostic protocols for badnaviruses has restricted its exploitation. Sukal et al. (2017) identified three DBVs, namely DBALV2, DBESV, and DBVRTV2, in the Pacific germplasm collection at CePaCT using a rolling circle amplification (RCA)-based strategy. In this study, RCA was used to identify and further characterize the molecular diversity of badnaviruses infecting the Pacific yam germplasm collection. The present study brings the number of Pacific DBV complete genome sequences to 24, suggesting geographic restriction of DBVs in the Pacific, a key consideration for future germplasm screening and dissemination.

2 | MATERIALS AND METHODS

2.1 | Samples and nucleic acid extraction

SPC-CePaCT holds 330 accessions in its in vitro germplasm collection. For this study, 224 out of the 330 yam accessions from the collection were acclimatized in SPC-CePaCT’s insect-proof greenhouse. After 3 months, leaf samples were collected and total nucleic acids (TNA) extracted (Kleinow et al., 2009). The purified TNA was quantified using a spectrophotometer (NanoDrop 2000; Thermo Fisher Scientific) and the concentration adjusted to c. 500 ng/μl with sterile nuclease-free water (NF-H2O). Episomal DBV-free yam accession DA/NGA01, a Nigerian accession obtained from the International Institute of Tropical Agriculture (IITA), was used as a virus-free control for the RCA experiments. This accession had previously tested negative for episomal DBV both at IITA using immunocapture (IC)-PCR and at SPC-CePaCT using IC-PCR and RCA according to protocols described in Seal et al. (2014). All yam accession descriptors included in the text refer to those used at SPC-CePaCT.

2.2 | RCA, cloning and Sanger sequencing

A badnavirus-biased RCA approach (Sukal et al., 2019) was used to amplify viral circular DNA. Briefly, a mixture of 32 degenerate badnavirus primers (Sukal et al., 2019) at a final concentration of 0.4 μM of each primer, 1 × Ph29 buffer (NEB) and 1 μl (c. 500 ng) of TNA was made up to a final volume of 10 μl with sterile NF-H2O and denatured at 95 °C for 3 min, cooled to 4 °C and placed on ice. A 10 μl reaction mixture consisting of 2.5 μM exo-resistant random hexamers (Thermo Fisher Scientific), 1 × Ph29 buffer, 2 ng/μl bovine serum albumin (BSA), 4 mM DTT, 15 mM dNTPs, 5 U/μl of Ph29 DNA polymerase (Thermo Fisher Scientific) and sterile NF-H2O to make up the final volume, was prepared and added to each denatured sample. Reactions were incubated at 36 °C for 18 hr, followed by 65 °C for 10 min to denature the Ph29 DNA polymerase.

RCA products were independently digested with EcoRI and SphI (NEB). These enzymes were selected following in silico restriction analysis of published yam badnavirus genome sequences and previously published work (Sukal et al., 2019). Digest products were electrophoresed through 1.5% agarose gels, stained with SYBR Safe (Thermo Fisher Scientific) and fragments of interest were excised, purified and cloned as described in Sukal et al. (2017). Sequencing was carried out using BadnaFP/RP primers (Yang et al., 2003) and the resulting reads were queried against GenBank using the BLASTn and BLASTx algorithms to determine identity. A primer-walking approach using standard Sanger sequencing conditions was subsequently used to sequence the complete genomes of four DBALV isolates, including three samples (one each of D. alata, D. transversa, and D. trifida) from Vanuatu and one D. esculenta sample from Tonga. The presence of the putative single SphI restriction site was confirmed with PCR using sequence-specific primers (DBALV-SphI site_Fwd1 5′-GCTGAGGTACTTTATCAC-3′ and DBALV-SphI site_Rev1 5′-GGTTGGGCGCTATGTA-3′) flanking the site as described previously (Sukal et al., 2017).

2.3 | Next-generation sequencing and genome assembly

Undigested RCA products of 10 Papua New Guinea (PNG) samples and 6 Vanuatu samples were purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) and sent to the Central Analytical Research Facility (CARF), Queensland University of Technology, Brisbane, Australia, for library preparation and sequencing using the Illumina MiSeq system to generate paired-end reads of 301 bp each. Similarly, undigested RCA products from a further 98 samples that did not produce any apparent visible restriction digest profiles following digestion of RCA products using EcoRI and SphI were purified and pooled by country and sent for
the preparation of 13 additional libraries. This included six libraries using samples from Fiji, two libraries each for samples from Vanuatu, New Caledonia and Federated States of Micronesia (FSM), and one library for samples from PNG, which were subsequently sequenced by next-generation sequencing (NGS) as described previously.

Quality of the raw reads was assessed with FastQC v. 0.10.1 (Babraham Bioinformatics). A pipeline similar to that described in Muller et al. (Babraham Bioinformatics). A pipeline similar to that described in Muller et al. (2015) for the processing of NGS data and characterization of badnavirus diversity. Raw reads were trimmed to obtain optimum quality using the dynamic trim function of SolexaQA++ v. 3.1.3 (Cox et al., 2010) and FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Quality-corrected reads were then mapped against the D. alata reference genome (GenBank accessions: CZHE02000001–CZHE0205770) using the Geneious v. 11.0.2 (http://www.geneious.com) reference mapper algorithm with default settings. Unmapped reads were assembled into contigs using the SPAdes v. 3.5.0 algorithm (Bankevich et al., 2012) with k-mers 21, 33 and 55 on the Galaxy platform (Afgan et al., 2015). The resulting contigs were imported into Geneious and BLASTn analysis was performed against a local database of the complete genome sequences of all known Caulimoviridae. Finally, the viral genome having the highest nucleotide identity with each assembled contig was used for reference-guided mapping to generate a consensus viral genome. Putative viral open reading frames (ORFs) and conserved badnaviral sequence motifs were identified using Geneious and ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/).

### 2.4 Pairwise sequence comparison, phylogenetic and recombination analysis

Partial coding sequences for reverse transcriptase (RT)-ribonuclease H (RNase H), delimited by the BadnaFP/RP primers, were used to determine the pairwise nucleotide sequence identity using the Sequence Demarcation Tool (SDT v. 1.2; Muhire et al., 2014). For phylogenetic analysis, multiple sequence alignments were constructed using ClustalW (Larkin et al., 2007) within MEGA 7 (Kumar et al., 2016) and the maximum-likelihood method (Kimura-2-parameter model) was used to reconstruct phylogenetic trees following 1,000 bootstrap iterations. To assess possible recombination events, sequences from the different DBALV and DBALV2 groups from this study together with representative isolates from the NCBI database were analysed using the RDP4 program (with embedded RDP, GENECONV, BOOTSCAN, MAXCHI, CHIMAERA, 3SEQ, and SISCAN tools) using the default parameters (Martin et al., 2015).

### 3 RESULTS

#### 3.1 RCA, Sanger sequencing and NGS

Total nucleic acids extracts from 224 yam samples representing 5 yam species, including D. alata (185), D. esculenta (31), D. bulbifera (6), and 1 each of D. transversa and D. trifida, were subjected to RCA, followed by independent restriction digestion using either EcoRI or Sphl. A total of 35 samples from PNG, Vanuatu, and Tonga produced restriction profiles indicative of the presence of badnaviruses (Table 1). In some instances, the total size of the EcoRI restriction digest products was larger than for a single badnavirus genome, suggesting mixed infections of different virus genotypes.

Of the 185 D. alata accessions tested, only 15/38 samples from PNG and 2/58 samples from Vanuatu produced restriction profiles indicative of badnavirus infection (Table 1). For the 15 PNG samples, 5 distinct EcoRI restriction profiles were observed in digests of RCA-amplified DNA (Figure 1a; Table 2). Using Sphl digestion of RCA-amplified DNA, a single c. 8 kb band was observed for 14 of the 15 PNG accessions, while accession DA/PNG58 produced 2 bands of c. 1 and 7 kb (Figure 1b). For the two positive D. alata Vanuatu samples (DA/VUT12 and DA/VUT38), EcoRI digestion of RCA-amplified DNA of DA/VUT12 and DA/VUT38 produced distinct restriction profiles (Figure 2a, lanes 1 and 2; Table 2), while digestion using Sphl produced either a single c. 8 kb band (Figure 2b, lane 1) or three bands of c. 2.5, 3.5 and 5 kb (Figure 2b, lane 2), respectively.

Of the six D. bulbifera accessions tested, all five from Vanuatu produced restriction profiles comprising six or seven fragments of less than 4 kb following digestion with EcoRI (Figure 2a, lanes 3–6; Table 2), and a single putative full-length digest fragment using Sphl (Figure 2b, lanes 3–6). The single accession tested from Samoa produced no visible digest fragments. Two distinct restriction profiles were observed following EcoRI digestion of the RCA-amplified DNA from D. bulbifera accessions from Vanuatu. Three samples (DB/VUT01, 02 and 03) produced a profile with seven fragments (Figure 2a, lanes 3–4), while digestion of two samples (DB/VUT04 and 07) produced six fragments (Figure 2a, lanes 5–6; Table 2).

Of the 31 D. esculenta accessions tested, 11 showed profiles indicative of badnavirus infection including the 2 samples tested from Tonga and all 9 samples tested from Vanuatu, while no restriction profiles were observed for samples from Fiji, PNG or Samoa. EcoRI digestion of RCA-amplified DNA from the nine positive D. esculenta samples from Vanuatu (DE/VUT01–09; Figure 2a, lanes 7–8) and two positive D. esculenta from Tonga (DE/TON02–03; Figure 2a, lanes 9–10) resulted in the same restriction profile as previously observed for accession DA/VUT12 (Figure 2a, lane 1), while digestion with Sphl produced a single c. 8 kb fragment for all 11 samples (Figure 2b, lanes 7–10).

When RCA-amplified DNA from the single D. transversa accession (DTV/VUT01) was digested using EcoRI (Figure 2a, lane 11), the profile matched that of the Vanuatu D. bulbifera samples DB/VUT04 and 07 (Figure 2a, lanes 5–6). Similarly, the restriction profile of the EcoRI-digested RCA-amplified DNA from the single D. trifida accession (DTF/VUT01) matched that of the D. alata accession DA/VUT12 as well as the D. esculenta accessions from Vanuatu and Tonga described previously (Figure 2a, lane 12). In both samples, digestion with Sphl produced a single c. 8 kb fragment.
For all 35 putative badnavirus-infected samples, the SphI-digested RCA products were cloned and sequenced using BadnaFP/RP primers. BLASTn analysis of the partial RT/RNase H-coding sequences obtained following sequencing with the BadnaFP/RP primers revealed that all 15 sequences originating from PNG showed 89%–100% nucleotide identity with DBALV2 isolate PNG10 (GenBank accession no. KY827395), while the 18 sequences from Vanuatu and 2 sequences from Tonga showed 85%–92% nucleotide identity with DBALV isolate 2ALa (GenBank accession no. KX008571).

Complete genome sequences were subsequently obtained using a primer-walking strategy from samples DA/VUT12, DTV/VUT01, and DTF/VUT01 from Vanuatu, as well as DE/TON02 from Tonga. The complete genome sizes of these four DBALV isolates comprised 7,531, 7,503, 7,579, and 7,390 bp, respectively, and the complete annotated sequences were submitted to GenBank under accession numbers MH404165–MH404168.

An additional 16 RCA-positive samples, including 10 D. alata samples from PNG and 6 samples from Vanuatu (1 D. alata, 2 D. esculenta, and 3 D. bulbifera) were subsequently chosen and undigested RCA products were sequenced using NGS. Using the NGS data from the 16 individual RCA-positive samples (Tables S1 and S2), complete badnavirus genomes were assembled. The genomes generated from all the PNG samples (Table S1) were most similar to DBALV2, while the genomes assembled from Vanuatu samples (Table S2) were all most similar to DBALV. The 16 complete genomes were annotated and submitted to GenBank under accession numbers MH404155–MH404164 and MH404169–MH404174. The NGS data obtained from the 13 pooled libraries resulted in contigs with no significant identity to viral sequences.

### Table 1: Summary of accessions in the SPC-CePaCT Yam collection available for testing in this study and the results of rolling circle amplification (RCA)

| Accession | Tested | Positive |
|-----------|--------|----------|
| Fiji      | 42     | 0        |
| FSM       | 21     | 0        |
| PNG       | 38     | 0        |
| New Caledonia | 24 | 0    |
| Samoa     | 2      | 1        |
| Tonga     | 0      | 0        |
| Vanuatu   | 58     | 2        |
| Total     | 185    | 17       |

Note: No. tested = number of accessions tested for badnavirus. No. positive = number of accessions considered positive for badnavirus following restriction profiling of RCA products.

3.2 | Dioscorea bacilliform AL virus from the Pacific

Using NGS, six complete DBALV genomes from Vanuatu were generated including one from D. alata, three from D. bulbifera and two from D. esculenta. In addition, four complete genomes were generated using Sanger sequencing, including one each from D. alata, D. transversa, and D. trifida from Vanuatu as well as one complete genome from D. esculenta originating from Tonga. The complete genomes ranged from 7,390 to 7,579 bp in length with GC contents varying between 42.8% and 43.7%. All contained three ORFs typical of badnaviruses with ORFs 1–3 comprising 432, 378, and 5,670–5,736 bp, respectively (Table S2). The single intergenic region (IR), which varied from 885 to 1,038 bp, contained several conserved motifs typical of plant double-stranded (ds)DNA viruses. The putative trRNA<sup>met</sup> binding site (5′-TGGTATCAGAGCTCGGTT-3′), which showed 88.9% nucleotide identity to the plant trRNA<sup>met</sup> consensus sequence (3′-ACCAUAGUCUCGGUCCAA-5′) and which has been described as the priming site for reverse transcription, was designated as the origin of the circular genomes consistent with the convention used for other badnaviruses. A TATA-box (TATATAA) and polyadenylation signal, analogous to the 35S promoter of cauliflower...
mosaic virus (CaMV), were also identified in the region 5′ of the tRNA<sup>met</sup> site in all genomes.

BLASTn analysis of the full-length genome sequences confirmed that all had highest nucleotide identity to DBALV (GenBank accession numbers KX008571, KX008572, and KX008573) with nucleotide sequence identity ranging from 89.2% to 90%. BLASTn analysis using the partial RT/RNase H-coding region of each sequence revealed 97.5%–99.8% nucleotide identity to partial RT/RNase H-coding sequences previously obtained from Vanuatu yams (GenBank accessions AM072705–AM072708). The partial RT/RNase H-coding sequence delineated by the BadnaFP/RP primers was identified in silico and used to carry out pairwise sequence comparison (PASC) and phylogenetic comparison amongst the sequences and with other published DBALV sequences. PASC revealed that DBALV sequences from Vanuatu and Tonga had 92%–100% nucleotide identity amongst each other and 85%–94% to previously published DBALV isolates from Africa (Figure 3a). Phylogenetic analysis using the partial RT/RNase H-coding region revealed that the DBALV isolates from Africa were ancestral to the isolates from the Pacific (Figure 3b).

Recombination analysis did not detect any potential recombination event(s) in any of the DBALV full-length sequences.

### 3.3 Dioscorea bacilliform AL virus 2

RCA combined with NGS was used to generate 10 complete genome sequences from <i>D. alata</i> accessions from PNG. The complete genomes ranged from 7,843 to 7,879 bp in length and had a GC content of 41%–42.8%. Consistent with the previously published DBALV2 isolate PNG10 (GenBank accession number KY827395), these sequences contained three ORFs with ORF1 varying in length from 429 to 447 bp, ORF2 varying in length from 396 to 399 bp and ORF3 varying in length from 5,811 to 5,856 bp (Table S1). Conserved motifs including the tRNA<sup>met</sup> site (5′-TGGTATCAGAGC-3′, underlined nucleotides were not conserved in all sequences), TATA-boxes (TATATAA) and polyadenylation signals were identified in the IR, similar to those present in the DBALV sequences.
TABLE 2 Summary of EcoRI restriction digest profiles of rolling circle-amplified samples positive for badnavirus from yams of Papua New Guinea (DA/PNG05, 06, 12, 14, 17, 20, 22, 23, 43, 45, 51, 57, 58, and 59), Vanuatu (DA/VUT12 and 38, DB/VUT01–03, 04 and 07, DE/VUT01–09, DTF/VUT01, and DTV/VUT01) and Tonga (DE/TON02 and 03)

| Virus       | Restriction profile (kb) | Corresponding figure | Lanes with representative samples                  | Positive samples                      |
|-------------|--------------------------|----------------------|----------------------------------------------------|---------------------------------------|
| DBALV2      | 0.3, 0.5, 0.6, 0.8, 1.0, 1.6, 3.0 | 1a                   | 1, 2                                              | DA/PNG06, 20, 22, 23, 43, 51, 57, 59   |
|             | 0.6, 0.8, 1.1, 1.6, 2.5, 3.0, 4.0 | 1a                   | 3, 4                                              | DA/PNG07, 45                          |
|             | 0.5, 0.8, 1.0, 1.6, 2.3, 3.0 | 1a                   | 5, 6                                              | DA/PNG05, 12, 14                       |
|             | 1.4, 2.5, 4.0           | 1a                   | 7                                                 | DA/PNG17                              |
|             | 0.6, 0.8, 2.5, 4.0      | 1a                   | 8                                                 | DA/PNG58                              |
| DBALV       | 1.0, 1.2, 1.7, 3.3      | 2a                   | 1, 7, 8, 12                                       | DA/VUT12, DE/VUT01–09, DE/TON02–03, DTF/VUT01 |
|             | 1.0, 1.2, 1.7, 3.5      | 2a                   | 2                                                 | DA/VUT38                              |
|             | 0.9, 1.0, 1.1, 1.2, 1.7, 3.0, 3.4 | 2a               | 3, 4                                              | DB/VUT01–03                           |
|             | 0.9, 1.0, 1.2, 1.7, 3.4, 4.0 | 2a                   | 5, 6, 11                                          | DB/VUT04, 07, DTV/VUT01               |

Abbreviations: DBALV, Dioscorea bacilliform AL virus; DBALV2, Dioscorea bacilliform AL virus 2.

BLASTn analysis of the full genome sequences revealed a sequence identity of 87%-89.1% to the previously published DBALV isolate PNG10, while BLASTn analysis using the partial RT/RNase H-coding region nucleotide sequences revealed 90.7%-97.5% nucleotide identity to partial RT/RNase H-coding sequences generated previously from PNG D. alata (GenBank accessions AM072674, AM072683, and AM072685). PASC and phylogenetic analysis of DBALV2 was carried out using the partial RT/RNase H-coding sequences as described earlier. PASC between the DBALV2 sequences identified in this study, together with the previously published complete DBALV2-PNG10 sequence from PNG, revealed 89%-100% nucleotide identity between the sequences (Figure 4a), with the DBALV2-PNG10 sequence appearing to be ancestral to the 10 sequences characterized in this study (Figure 4b). No potential recombination event(s) were detected in any of the the full-length DBALV2 sequences.

4 | DISCUSSION

In this study, DBVs present in the Pacific yam germplasm collections held in trust by SPC-CePaCT were identified and characterized. Of the 224 accessions tested using RCA, 35 samples gave restriction profiles suggestive of episomal badnavirus infection. Based on differences in restriction profiles, partial RT/RNase H-coding nucleotide sequences, host species or the country of origin, a total of 20 complete genome sequences were generated and were identified as either DBALV or DBALV2. Importantly, in 98 samples that were considered negative using RCA, NGS of undigested RCA products failed to identify any sequences related to viruses, suggesting that the RCA protocol successfully identified all DBV-infected samples.

Based on this, and other studies (Kenyon et al., 2008), DBALV appears to be the most prevalent badnavirus in Pacific yam germplasm, with the virus identified in five yam species (D. alata, D. bulbifera, D. esculenta, D. transversa, and D. trifida) from Vanuatu as well as D. esculenta from Tonga. Using a PCR-based approach, Kenyon et al. (2008) reported the presence of DBALV in both Vanuatu and Solomon Islands. Unfortunately, samples from Solomon Islands were not available during the present study. Whereas a very low prevalence (3.4%) of DBALV was found in D. alata accessions from Vanuatu, all D. esculenta and D. bulbifera samples tested from Vanuatu were found to be infected in this study. The low prevalence of DBALV in Vanuatu D. alata could be because 31 of the 73 accessions in the Pacific collection were obtained from Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD). These accessions had previously been tested at CIRAD for viruses, including DBV, and only virus-free accessions were provided to SPC-CePaCT for inclusion in the Pacific yam collection (Filloux and Girard, 2006). The 31 accessions obtained from CIRAD were also found to be DBV-negative in this study. Previously, Filloux and Girard (2006) reported a low incidence of DBALV in Vanuatu D. alata, with only 4 of 56 accessions testing positive. An additional 42 previously untested Vanuatu D. alata accessions screened in this study were also found to have a low prevalence of DBALV, with only 2 samples testing positive. Additionally, no samples from Fiji, FSM, PNG, New Caledonia or Samoa tested positive for DBALV in the present study.

DBALV2 was only detected in D. alata samples from PNG, with an incidence of 42.5%, which was significantly higher than the incidence of DBALV infection of D. alata from Vanuatu. Kenyon et al. (2008) also reported DBALV2 from D. alata accessions only; however, they detected DBALV2 in samples from both PNG and Vanuatu. In contrast, DBALV2 was not detected in the samples from Vanuatu or from other countries included in this study.

Although DBESV has been previously reported and characterized from a D. esculenta accession from Fiji (DE/FJ14; Sukal et al., 2017), this virus was not detected in any additional D. esculenta accessions, or accessions of other yam species from either Fiji or any other countries included in the present study. Previously, Kenyon
FIGURE 2  (a) EcoRI and (b) SphI restriction analysis of representative samples of rolling circle amplified DNA from yams of Vanuatu (VUT) and Tonga (TON), showing profiles positive for the presence of badnaviruses. Lanes 1–12 are samples DA/VUT12, DA/VUT38, DB/VUT01, DB/VUT02, DB/VUT04, DB/VUT07, DE/VUT01, DE/VUT02, DE/TON02, DE/TON03, DTV/VUT01, and DTF/VUT01. Lane 13 is a known negative sample (DA/NGA01) and lane 14 is a no template control. M, GeneRuler 1 kb DNA ladder; DA, Dioscorea alata; DB, D. bulbifera; DE, D. esculenta; DTV, D. transversa; DTF, D. trifida

FIGURE 3  (a) Pairwise sequence comparison and (b) phylogenetic analysis (ClustalW alignment using the maximum-likelihood method [Kimura-2-parameter model]) of partial RT/RNase H-coding nucleotide sequences, showing the relationships of Dioscorea bacilliform AL virus (DBALV) isolates from this study with previously published complete sequences DBALV-[2ALa], DBALV-[2ALb] and DBALV-[3RT] from Nigeria (Bömer et al., 2016). DBALV2-PNG10 (KY827395) was used as the outgroup for the analysis. [Colour figure can be viewed at wileyonlinelibrary.com]
et al. (2008) reported DBESV-like RT/RNase H-coding partial sequences from D. alata accessions originating from Fiji, PNG, Tonga, and Vanuatu, as well as D. esculenta accessions originating from Fiji, PNG, and Solomon Islands. Because the focus of the present study was to characterize the episomal badnavirus sequences present in yam, the presence of endogenous DBESV was not addressed here but is worthy of future investigation.

DBRTV2 has been previously characterized from a single Samoan D. rotundata accession (Sukal et al., 2017), while Bömer et al. (2016) also detected and characterized DBRTV2 from D. rotundata originating from Nigeria. Because the SPC-CePaCT collections do not include D. rotundata from any additional Pacific countries, the presence of DBRTV2 in countries other than Samoa could not be ascertained. Further field collections and testing should be done to determine whether DBRTV2 is more widespread in the Pacific and whether it occurs in yam species other than D. rotundata. Because D. rotundata originates from West Africa (Lebot, 2009), it is possible that DBRTV2 may also originate from West Africa and was distributed into the Pacific with the introduction of D. rotundata accessions.

During restriction analysis of RCA products, the restriction profiles generated using EcoRI were found to vary considerably between virus isolates in the different yam species studied. Furthermore, in some instances, the size of EcoRI restriction fragments totalled greater than 8 kb. Possible explanations for this could be the presence of multiple viral sequences, the existence of sequence variants within an accession, or co-amplification of plant genomic DNA during RCA or partial digestion of the DNA. Based on sequence analysis, which confirmed the presence of single DBV infections and the absence of host genomic DNA, it is likely that, in some samples, DBVs are probably present as a collection of sequence variants or quasispecies that produce slightly variable digest patterns using some enzymes, such as EcoRI. Therefore, it may be necessary to use a combination of enzymes with single or multiple recognition sites within the virus genome for diagnostic purposes. It was also found that other enzymes, such as KpnI and Stul, can complement EcoRI and SphI in determining a DBV-positive sample (data not shown).

This study significantly increases the knowledge of yam-infecting badnaviruses in the Pacific, with 24 complete genome sequences belonging to four different species, infecting six species of yam from five countries, now available. Studies on Pacific yams show that the dominant yam badnaviruses present in Pacific germplasm differ from that of the African region. Although some virus species, such as DBALV, are present in both regions, several viruses identified in the African region have not been detected in the Pacific and vice versa. The present study shows that some of these viruses are restricted to only one or a few countries in the Pacific and, as such, special considerations must be taken to ensure that germplasm collections are thoroughly screened to prevent the dissemination of these badnavirus species to other countries. The availability of virus-tested yam germplasm is essential for the effective distribution and eventual use of yams for improved food and nutritional security in the Pacific. Testing of the Pacific yam germplasm for other viruses will also be necessary, however, before material can be made available for distribution.

**ACKNOWLEDGEMENTS**

The authors would like to thank the SPC-CePaCT for making their yam collections available for this project. Authors would also like to thank Dr Michael Furlong and Dr Grahame Jackson for their support and advice on this research. The funding for the project was provided by the Australian Centre for International Agricultural Research (#PC/2010/065). A.S. is a John Allwright Fellowship recipient.
CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID
Amit C. Sukal https://orcid.org/0000-0003-1634-2331
Rob M. Harding https://orcid.org/0000-0002-9600-7287

REFERENCES
Adams, M.J. and Carstens, E.B. (2012) Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2012). Archives of Virology, 157, 1411-1422.

Adams, M.J., Lefkowitz, E.J., King, A.M.Q., Harrach, B., Harrison, R.L., Knowles, N.J. et al. (2018) Changes to taxonomy and the international code of virus classification and nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). Archives of Virology, 162, 2505-2538.

Afgan, E., Soggetti, C., Goonasekera, N., Makunin, I., Benson, D., Crowe, M. et al. (2015) Genomics virtual laboratory: a practical bioinformatics workbench for the cloud. PLoS ONE, 10, e0140829.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S. et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology, 19, 455–477.

Bömer, M., Turaki, A., Silva, G., Kumar, P. and Seal, S. (2016) A sequence-independent strategy for amplification and characterization of episomal badnavirus sequences reveals three previously uncharacterized yam badnaviruses. Viruses, 8, 188.

Bömer, M., Rathnayake, A.J., Visendi, P., Silva, G. and Seal, S.E. (2018) Complete genome sequence of a new member of the genus Badnavirus, Dioscorea bacilliform RT virus 3, reveals the first evidence of recombination in yam badnaviruses. Archives of Virology, 163, 533–538.

Bömer, M., Rathnayake, A.J., Visendi, P., Sewe, S.O., Sicat, J.P.A., Silva, G. et al. (2019) Tissue culture and next-generation sequencing: a combined approach for detecting yam (Dioscorea spp.) viruses. Physiological and Molecular Plant Pathology, 105, 54–66.

Bourke, R.M. and Vlassak, V. (2004) Estimates of Crop Food Production in Papua New Guinea. Canberra, Australia: Australian National University.

Boukemel, M., Durand, O., Scarcelli, N., Lebas, B.S.M., Kenyon, L., Marchand, J.-L. et al. (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.

Cox, M.P., Peterson, D.A. and Biggs, P.J. (2010) SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics, 11, 485.

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

FAOSTAT. (2018) Production statistics. Available at: http://www.fao.org/faostat/en/#data [Accessed 29 November 2019].

Filloux, D. and Girard, J. (2006) Indexing and elimination of viruses infecting yams (Dioscorea spp.) for the safe movement of germplasm. In: Proceedings of the 14th Triennial Symposium of the ISRTC. Trivandrum, India: ISTRC. Available at: https://agritrop.cirad.fr/540794/1/document_540794.pdf [Accessed 29 November 2019].

Kenyon, L., Lebas, B.S.M. and 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.

Kleinow, T., Nischang, M., Beck, A., Kratzer, U., Tanwir, F., Preis, W. et al. (2009) Three C-terminal phosphorylation sites in the abutilon mosaic virus movement protein affect symptom development and viral DNA accumulation. Virology, 390, 89–101.

Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution, 33, 1870–1874.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics, 23, 2947–2948.

Lebot, V. (2009) Tropical Root and Tuber Crops: Cassava, Sweet Potato, Yams and Aroids. Wallingford: CABl.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A. and Muhire, B. (2015) RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evolution, 1, 1–5.

Muhire, B.M., Warsami, A. and Martin, D.P. (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PLoS ONE, 9, e108277.

Muller, E., Ravel, S., Agret, C., Abrokwa, F., Dzahini-Obiaye, H., Galyoun, I. et al. (2018) Next generation sequencing elucidates cacao badnavirus diversity and reveals the existence of more than ten viral species. Virus Research, 244, 235–251.

Odu, B.O., Hughes, J.D.A., Asiedu, R., Ng, N.Q., Shoyinka, S.A. and Oladiran, O.A. (2004) Responses of white yam (Dioscorea rotundata) cultivars to inoculation with three viruses. Plant Pathology, 53, 141–147.

Phillips, S., Briddon, R.W., Brun, A.A. and Hull, R. (1999) The partial characterization of a badnavirus infecting the greater Asiatic or water yam (Dioscorea alata). Journal of Phytopathology, 147, 265–269.

Seal, S., Turaki, A., Muller, E., Kumar, P.L., Kenyon, L., Filloux, D. et al. (2014) The prevalence of badnaviruses in West African yams (Dioscorea cayennensis-rotundata) and evidence of endogenous pararetrovirus sequences in their genomes. Virus Research, 186, 144–154.

Silva, G., Bömer, M., Rathnayake, A.J., Sewe, S.O., Visendi, P., Oyekanmi, J.O. et al. (2019) Molecular characterization of a new virus species identified in yam (Dioscorea spp.) by high-throughput sequencing. Plants, 8, 167.

Sukal, A.C., Taylor, M. and Tula, V.S. (2015) Viruses and their impact on the utilization of plant genetic resources in the Pacific. Acta Horticulturae, 1101, 127–132.

Sukal, A.C., Kidanemariam, D., Dale, J., James, A. and Harding, R. (2017) Characterization of badnaviruses infecting Dioscorea spp. in the Pacific reveals two putative novel species and the first report of Dioscorea bacilliform RT virus 2. Virus Research, 238, 29–34.

Sukal, A.C., Kidanemariam, D.B., Dale, J.L., Harding, R.M. and James, A.P. (2019) Assessment and optimization of rolling circle amplification protocols for the detection and characterization of badnaviruses. Virology, 529, 73–80.

Yang, I.C., Hafner, G.J., Dale, J.L. and Harding, R.M. (2003) Genomic characterization of taro bacilliform virus. Archives of Virology, 148, 937–949.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Sukal AC, Kidanemariam DB, Dale JL, Harding RM, James AP. Characterization and genetic diversity of Dioscorea bacilliform viruses present in a Pacific yam germplasm collection. Plant Pathol. 2020;69:576–584. https://doi.org/10.1002/ppa.13133