Arch Virol (2017) 162:517–521
DOI 10.1007/s00705-016-3113-3

BRIEF REPORT

The genome sequence of Dioscorea bacilliform TR virus, a member of the genus *Badnavirus* infecting *Dioscorea* spp., sheds light on the possible function of endogenous *Dioscorea* bacilliform viruses

Marie Umber[1] · Rose-Marie Gomez[1] · Suzia Gélabale[1] · Lydiane Bonheur[1,2] · Claudie Pavis[1] · Pierre-Yves Teycheney[2]

Received: 20 August 2016 / Accepted: 5 October 2016 / Published online: 21 October 2016

Abstract The complete genome sequence of Dioscorea bacilliform TR virus (DBTRV) was determined. The closest relatives of DBTRV are Dioscorea bacilliform AL virus (DBALV) and Dioscorea bacilliform RT virus 1 (DBRTV1). Specific primers were designed and used to determine the prevalence of DBTRV in a yam germplasm collection. It was found that this virus infects *Dioscorea alata* and *D. trifida* plants in Guadeloupe and French Guyana. DTRBV was not detected in any of the tested *D. cayenensis-rotundata* accessions. *In silico* analysis provided evidence for the presence of DBTRV-like endogenous sequences in the genome of *D. cayenensis-rotundata*, pointing to a possible role of these sequences in antiviral defense.

Keywords *Dioscorea* · Badnavirus · Yam · Dioscorea bacilliform TR virus · Genome

The nucleotide sequences reported in this work have been deposited in the GenBank database under accession numbers KX430225, KX430226, KX430229, KX430235 to KX430252 and KX430257.

Electronic supplementary material The online version of this article (doi:10.1007/s00705-016-3113-3) contains supplementary material, which is available to authorized users.

✉ Pierre-Yves Teycheney
teycheney@cirad.fr

Marie Umber
marie.umber@antilles.inra.fr

[1] INRA, UR ASTRO, 97170 Petit-Bourg (Guadeloupe), France

[2] CIRAD, UMR AGAP,
97130 Capesterre-Belle Eau (Guadeloupe), France

Yams (*Dioscorea* spp.) are an important staple food worldwide, particularly in West Africa and the South Pacific [1, 2], and also in the Caribbean. They play a key role throughout the intertropical zone in agricultural diversification programs aimed at increasing food self-sufficiency. Within the last 10 years, several viruses of the genera *Aureusvirus*, *Badnavirus*, *Carlavirus, Comovirus*, *Cucumovirus*, *Fabavirus*, *Macluravirus*, *Potexvirus* and *Potyvirus* have been reported in cultivated yams, revealing the wide range of viruses infecting this crop. The fact that cultivated yams are vegetatively propagated is likely to promote the accumulation of viruses to the point that multiple infections are common, making it difficult to assess the symptoms and impact on yields of individual viruses.

Badnaviruses (family *Caulimoviridae*, genus *Badnavirus*) are studied extensively because of their high prevalence and economic impact on important tropical crops such as cocoa, banana, sugarcane, citrus and yams and their high molecular diversity [3]. The genomes of four distinct badnaviruses infecting yams have been completely sequenced: Dioscorea bacilliform AL virus (DBALV), Dioscorea SN bacilliform virus (DBSNV), Dioscorea bacilliform RT virus 1 (DBRTV1) and Dioscorea bacilliform RT virus 2 (DBRTV2) [4–6]. These sequences were used for phylogenetic studies, together with partial nucleotide sequences generated by PCR using the badnavirus-specific primer pair BadnaFP/BadnaRP targeting the conserved reverse transcriptase/ribonuclease H (RT/RNaseH) domain [7]. These analyses revealed the existence of 15 distinct badnavirus types in yams [6–9]. However, apart from DBALV, DBSNV, DBRTV1 and DBRTV2, evidence is still lacking that episomal forms of these viruses are still extant. They might therefore exist as endogenous forms called endogenous *Dioscorea* bacilliform viruses (eDBVs) [10].

A search for badnaviruses was carried out by PCR in 65 *D. trifida* accessions maintained in the collection of the Biological Resource Center for Tropical Plants (BRC-TP) of the French West Indies. For this, total genomic DNA was extracted from vitroplants using a DNeasy Plant Mini Kit (QIAGEN, Courtaboeuf, France). DNA quality was assessed by PCR using the atpB1/atpB2 primer pair and conditions described by Soltis *et al.* [11]. Badnavirus-specific PCR was performed using the BadnaFP/BadnaRP primer as described by Umber *et al.* [10]. An amplification product of the expected size (579 bp) was obtained from 96.9 % (63/65) of the analyzed samples (Table S1), suggesting that these PCR products were amplified from both episomal badnavirus genomes and eDBVs. A specific search for episomal forms was carried out using immunocapture-PCR (IC-PCR) performed on leaf extracts of the same 65 *D. trifida* accessions, as described by Umber *et al.* [12] with the following modifications: primer pair BadnaFP/BadnaRP was used for the amplification of badnavirus sequences, and primer pair atpB1/atpB2 was used for monitoring contamination by yam genomic DNA that could lead to false positives resulting from the amplification of eDBVs. A total of 38 samples (58.5 %) were found positive (Table S1), a steep decrease from the above-mentioned 96.9 % PCR positive samples, indicating that episomal badnaviruses were present in the analyzed *D. trifida* accessions and that a large proportion of the products obtained by PCR resulted from the amplification of eDBVs.

A selection of 11 IC-PCR products were cloned into pGEM-T Easy Vector (Promega, Charbonnières, France). One to four clones were sequenced for each amplification product (Beckman Coulter, Takeley, UK), resulting in a total of 21 sequences (Table S1). Phylogenetic analysis showed that all 21 sequences belonged to the yam badnavirus groups 8 and 9 defined by Kenyon *et al.* [8] and named DBV-A(A) and DBV-B by Bousalem *et al.* [9]. Interestingly, accessions #50 and #52 hosted sequences from both groups (Fig. 1A), illustrating that mixed infections by members of different badnavirus species occur in yams.

It is established that episomal forms of badnavirus from 4 of the 15 groups exist, since the genomes of members of these groups were fully sequenced from purified virus preparations (groups 4 and 8, [4, 5]) or following rolling-circle amplification (RCA), which selectively amplifies episomal forms of circular DNA viral genomes (groups 13 and 14, [6]). In contrast, nucleotide sequences from yam badnavirus group 9 were only reported as endogenous forms in *D. cayenensis-rotundata* genomes prior to this work [10, 13]. Therefore, the sequencing of the full-length episomal genome of a yam badnavirus from group 9 was undertaken. For this, a pair of outward-facing primers (DBV-BgenF/DBV-BgenR; see Fig. 2 and Table S2) was designed from sequence 7P_12 originating from IC-PCR performed on sample 496_Borelli (*D. trifida* accession id PT-IG-00432). We used 100 ng of genomic DNA from this sample as template in a long-PCR experiment. The PCR mix contained 400 nM of each primer, 400 µM of dNTP and 2.5 U of LongAmp *Taq* DNA Polymerase (New England Biolabs, Evry, France). PCR conditions were 4 mn at 94 °C; 25 cycles of 30 s at 94 °C, 30 s at 58 °C, 7 mn at 72 °C; and a final elongation step of 10 mn at 72 °C. A single PCR product of the expected size (7-8 kbp) was generated and cloned into pGEM-T Easy Vector. The insert of one recombinant clone (#496-8) was fully sequenced on both strands using a genome-walking strategy (Beckman Coulter), generating a 6685-bp sequence. *In silico* analysis showed that this sequence covered nearly the entire viral genome of a badnavirus (Fig. 2), lacking the 5' end of the intergenic region and the beginning of ORF1 and displaying three frameshifts resulting in disruptions of the reading frame. Additional primer pairs ORF2496F/ORF2496R, MID1ORF3F/MID1ORF3R, MID2ORF3F/MID2ORF3R, ENDORF3F/ENDORF3R and IG496F/IG496R (Table S2) were designed and used in PCR experiments on 50 ng of genomic DNA from sample 496_Borelli to generate amplification products covering sequence areas with ambiguities and to complete the sequence (Fig. 2). The reaction mixtures contained 200 nM of each primer and 1 U of Phusion DNA Polymerase (New England Biolabs). PCR conditions were 30 s at 98 °C; 30 cycles of 10 s at 98 °C, 30 s at 58 °C, and 45 s at 72 °C; and a final elongation step of 10 mn at 72 °C. PCR products were cloned in pGEM-T Easy Vector, and five recombinant clones per PCR product were sequenced on both strands. All of the sequences displayed >99 % identity to overlapping regions of clone #496-8. Sequence assembly was performed using Geneious R9 (Biomatters, Auckland, New Zealand) and resulted in a 7333-bp sequence (GenBank accession number KX430257) displaying all of the hallmarks of the genome of a badnavirus [14].

Three open reading frames were identified (Fig. 2). Like those of banana streak IM virus [15], ORFs 2 and 3 were in a -1 translational frame relative to the preceding ORF, the start and stop codons of ORFs 1 and 2 overlapped, and those of ORFs 2 and 3 were separated by two nucleotides. A putative negative-strand primer-binding site was identified (5'-TGGTATCAGAGCTTGGT-3'), and following convention, the 5' nucleotide of this motif was designated the beginning of the genome. Consistent with a leaky scanning model of translation, there were no internal AUG codons within either ORF1 or 2. Putative promoters were identified in the region of the genome between the end of ORF3 and the negative-strand primer-binding site and

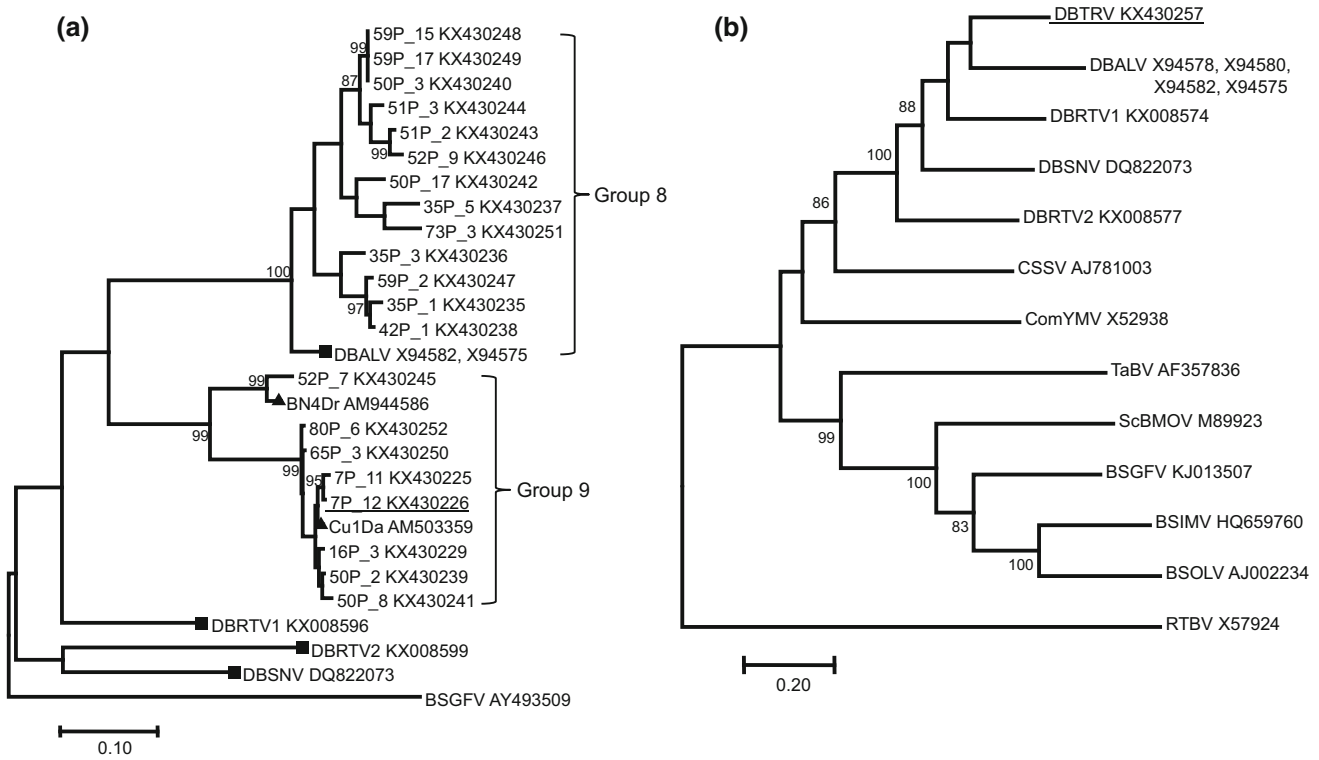

Fig. 1 Phylogenetic analysis of badnaviruses. Alignments were performed using CLUSTAW [19], and the evolutionary relationships were inferred using the maximum-likelihood method based on the HKY model [20] in MEGA7 [21]. Bootstrap values of 1,000 replicates are given above nodes when above 80 %. The scale bar shows the number of substitutions per base position. (a) Phylogenetic tree constructed from the 21 partial nucleotide sequences of the badnavirus RT/RNaseH domain amplified by IC-PCR from *Dioscorea trifida* accessions of the BRC-TP collection using the BadnaFP/BadnaRP primer pair. Additional sequences of the RT/RNaseH domain from other badnaviruses were used: Cu1Da and BN4Dr, which belong to badnavirus group 9, are indicated by dark triangles; Dioscorea bacilliform AL virus (DBALV), Dioscorea bacilliform SN virus (DBSNV), Dioscorea bacilliform RT virus 1 (DBRTV1) and Dioscorea bacilliform RT virus 2 (DBRTV2) are indicated by dark squares. Banana streak GF virus (BSGFV) was used as an outgroup. GenBank accession numbers of the sequences used are provided. Sequence 7P_12, from which outward-facing primers were designed for the amplification of the DBTRV genome is underlined. (b) Phylogenetic tree built from the complete nucleotide sequences of DBTRV and badnavirus type members cacao swollen shoot virus (CSSV), commelina yellow mottle virus (CoYMV), sugarcane bacilliform Mor virus (ScBMoV), banana streak GF virus (BSGFV), banana streak OL virus (BSOLV), banana streak IM virus (BSIMV) and taro bacilliform virus (TaBV). Rice tungro bacilliform virus (RTBV) was used as an outgroup. GenBank accession numbers of the sequences used are provided. DBTRV is underlined

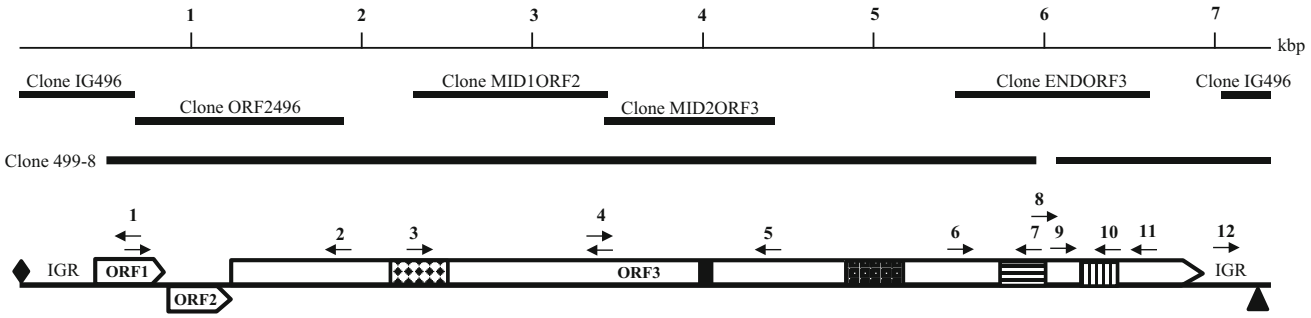

Fig. 2 Genome organization of DBTRV. A linearized genome map of Dioscorea bacilliform TR virus is shown with scale and clone positions indicated. Arrowed boxes mark open reading frames. Conserved protein domains as listed in the Pfam database (http://pfam.sanger.ac.uk/) are as follows: diamond hatching represents the viral movement protein domain (PF01107), black shading represents the zinc finger, black shading with white dots represents the retropepsin (pepsin-like aspartic protease) domain (CD00303), horizontal hatching represents the reverse transcriptase domain (CD01647), and vertical hatching represents the ribonuclease H1 domain (CD06222). The black diamond at the beginning and the black triangle at the end of the genome mark the position of the negative-strand primer-binding site and the TATA box, respectively. The position and orientation of the primers used in this study (see Table S2) is provided above the genome map

included a TATA-box (AGCTATATAAGCA) at position 7192-7204 and a CAAT-box (ACTCAATTATT) at position 7165-7175.

Conceptual translation of ORFs 1, 2 and 3 gave rise to proteins with molecular weights of 16.7, 13.6 and 215.4 kDa, respectively. Several conserved domains were detected in the polyprotein encoded by ORF3, including a dUTPase motif characteristic of the movement proteins of members of the family *Caulimoviridae* [16] at position 344-473 and domains associated with aspartic protease and reverse transcriptase/RNase H activities at positions 1196-1308 and 1457-1619, respectively (Fig. 2). A zing finger motif (CKCFLCGAEGHFARECPN) typical of the coat protein of members of the family *Caulimoviridae* [14] was found at position 920-937.

Sequence comparisons performed on the nucleotide sequence of the RT/RNaseH domain of the assembled genome sequence (position 5881-6408) showed that this sequence displayed 65.5 % to 73.7 % nucleotide sequence identity to the corresponding region of other fully sequenced episomal yam badnaviruses (Table 1). Considering that the species demarcation criterion for the genus *Badnavirus* is 80 % identity in this domain, the assembled genome sequence belongs to a new species within this genus, and we propose the name *Dioscorea* bacilliform TR virus (DBTRV) for this virus. Additional sequence comparisons performed on the nucleotide sequences of ORFs 1, 2 and 3 and the corresponding encoded amino acid sequences confirmed that DBTRV is a distinct yam-infecting badnavirus (Table 1). As shown in Fig. 1B, DBTRV and DBALV represent sister taxa in the genus *Badnavirus* and form a well-supported monophyletic group with other yam badnaviruses.

Taking advantage of a bacterial artificial chromosomes (BAC) library of *D. trifida* accession 519_Lac bleu 3 (accession id PT-IG-00455; H. Berges, INRA, Toulouse, France, unpublished), a search for endogenous sequences from yam badnavirus group 9, referred to as eDBV9, in the genome of this accession was attempted. For this, molecular hybridization was performed on BAC clones as described by Gayral *et al.* [17], using five probes covering DBTRV genome positions 531-849, 3404-4005, 4186-4783, 5855-6432 and 7022-289, respectively. No positive hybridization signal was obtained, suggesting that this *D. trifida* accession may not host DBTRV-like sequences (data not shown). Likewise, no eDBTRV-like sequence could be found in *D. alata* EST sequences available in GenBank following BLAST analysis (data not shown), suggesting that such sequences are not present in the *D. alata* genome either.

The prevalence of DBTRV was assessed in the yam germplasm *in vitro* collection maintained at the BRC-TP of the French West Indies. For this, a DBTRV-specific primer pair (DBTRV-F/DBTRV-R; Table S2) targeting DBTRV RT/RNaseH domain was designed based on the alignment of nucleotide sequences shown in Fig. 1A. A total of 253 accessions were screened by IC-PCR (see Table S1), using these primers and the conditions described above. Additional indexings were performed on the same leaf extracts using badnavirus-specific degenerate primer pair BadnaFP/BadnaRP. Indexings showed that 35.6 % (90/253) of the analyzed samples were infected by badnaviruses, that DBTRV alone was present in 14.6 % (37/253) of these samples and in 24.6 % (16/65) of the *D. trifida* samples (Table S3). Several of the DBTRV-infected *D. trifida* accessions originated from Guadeloupe and French Guyana and were conserved exclusively *in vitro* since their introduction in the BRC-PT collection, and therefore, the distribution area of DBTRV includes at least Guadeloupe and French Guyana. The other accessions used in this work were conserved under open-field conditions in Guadeloupe following their introduction and prior to their conservation as vitroplants, making it impossible to establish the geographical origin of DBTRV in these accessions. However, the design of specific detection tools now makes it possible to investigate the presence of DBTRV in all yam-growing areas. Interestingly, DBTRV was also detected in 16.9 % (21/124) of the indexed of *D. alata* accessions but not in any of the 34.6 % (18/52) badnavirus-infected *D. cayenensis-rotundata* accessions.

It is expected that access to the first complete genome sequence of a yam badnavirus from group 9 reported here could help refine the structure of rearranged eDBVs. Indeed, *in silico* analysis showed that uncharacterized parts of rearranged eDBV9 sequences recently described in the genome of *Dioscorea cayenensis-rotundata* [10] display up to 93 % identity with the DBTRV intergenic region (Table S4). Furthermore, several eDBV9 sequences

Table 1 Percentages of nucleotide and amino acid sequence identity between DBTRV and DBALV, DBSNV, DBRTV1 and DBRTV2. Figures for amino acid sequences are shown in parentheses

	Complete genome	RT-RNAseH domain	ORF1	ORF2	ORF3
DBALV	65.6 %	70.6 % (75.6 %)	72.5 % (81.1 %)	62.1 % (59.5 %)	64.7 % (67.2 %)
DBSNV	61.9 %	67.0 % (73.3 %)	72.0 % (79.7 %)	55.9 % (53.7 %)	62.9 % (64.8 %)
DBRTV1	64.4 %	73.7 % (83.0 %)	67.6 % (69.7 %)	60.5 % (61.2 %)	69.8 % (73.8 %)
DBRTV2	57.2 %	65.5 % (77.3 %)	61.2 % (54.9 %)	48.5 % (46.3 %)	60.6 % (62.8 %)

displayed 100 % identity to DBTRV over more than 25-nt-long sequence stretches, potentially allowing the synthesis of small interfering RNAs targeting DBTRV. It may therefore be hypothesized that these eDBV9 sequences could trigger a silencing-based defense against DBTRV, as hypothesized previously for other endogenous viral elements [18], and explain the absence of DBTRV-infected plants among *D. cayenensis-rotundata* accessions studied in this work.

Acknowledgements The authors wish to thank Laura Pazzé for technical help. This work was funded by INRA, Région Guadeloupe and the European Regional Development Fund.

Compliance with ethical standards

Ethical standard statement This study contained no experiments involving human participants or animals.

Conflict of interest Authors declare they have no conflict of interest.

Author consent The authors declare their consent to the content of the manuscript.

References

1. FAO Statistics (2014) Rome, Italy. http://www.faostat.fao.org
2. Maroya N, Asiedu R, Kumar PL, MignounaD Lopez-Montes A, Kleih U, Phillips D, Ndiame F, Ikeorgu J, Otoo E (2014) Yam improvement for income and food security in West Africa: effectiveness of a multi-disciplinary and multi-institutional teamwork. J Root Crops 40:85–92
3. Bhat AI, Hohn T, Selvarajan R (2016) Badnaviruses: the current global scenario. Viruses 8:177
4. Briddon RW, 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
5. Seal SE, Muller E (2007) Molecular analysis of a full-length sequence of a new yam badnavirus from *Dioscorea sansibarensis*. Arch Virol 152:819–825
6. Bömer M, Turaki AA, Silva G, Kumar PL, Seal SE (2016) A sequence-independent strategy for amplification and characterisation of episomal badnavirus sequences reveals three previously uncharacterised yam badnaviruses. Viruses 8:E188
7. Yang IC, Hafner GJ, Revill PA, Dale JL, Harding RM (2003) Sequence diversity of South Pacific isolates of Taro bacilliform virus and the development of a PCR-based diagnostic test. Arch Virol 148:1957–1968
8. Kenyon L, Lebas BSM, Seal SE (2008) Yams (*Dioscorea* spp.) from the South Pacific Islands contain many novel badnaviruses: implications for international movement of yam germplasm. Arch Virol 153:877–889
9. Bousalem M, Durand O, Scarcelli N, Lebas BSM, Kenyon L, Marchand JL, Lefort F, Seal SE (2009) Dilemmas caused by endogenous pararetroviruses regarding the taxonomy and diagnosis of yam (*Dioscorea* spp.) badnaviruses: analyses to support safe germplasm movement. Arch Virol 154:297–314
10. Umber M, Filloux D, Muller E, Laboureau N, Galzi S, Roumagnac P, Iskra-Caruana ML, Pavis C, Teycheney PY, Seal SE (2014) The genome of African yam (*Dioscorea cayenensis–rotundata* complex) hosts endogenous sequences from four distinct badnavirus species. Mol Plant Pathol 15(8):790–801
11. Soltis PS, Soltis DE, Chase MW (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature 402:402–404
12. Umber M, Pichaut JP, Farinas B, Laboureau N, Janzac B, Plaisir-Pineau K, Pressat G, Baurens FC, Chabannes M, Duroy PO, Guiougou C, Delos JM, Jenny C, Iskra-Caruana ML, Salmon F, Teycheney PY (2016) Marker-assisted breeding of *Musa balbisiana* genitors devoid of infectious endogenous Banana streak virus sequences. Mol Breeding 36(6):1–11
13. Seal S, Turaki A, Muller E, Kumar PL, Kenyon L, Filloux D, Galzi S, Lopez-Montes A, Iskra-Caruana ML (2014) The prevalence of badnaviruses in West African yams (*Dioscorea cayenensis–rotundata*) and evidence of endogenous pararetrovirus sequences in their genomes. Virus Res 186:144–154
14. King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (2012) Virus taxonomy. Ninth report of the international committee on taxonomy of viruses. Elsevier, San Diego
15. Geering AD, Parry JN, Thomas JE (2011) Complete genome sequence of a novel badnavirus, banana streak IM virus. Arch Virol 156(4):733–737
16. Hany U, Adams IP, Glover R, Bhat AI, Boonham N (2014) The complete genome sequence of Piper yellow mottle virus (PYMoV). Arch Virol 159(2):385–388
17. Gayral P, Noa-Carrazana JC, Lescot M, Lheureux F, Lockhart BEL, Matsumoto T, Piffanelli P, Iskra-Caruana ML (2008) A single banana streak virus integration event in the banana genome as the origin of infectious endogenous pararetrovirus. J Virol 82:6697–6710
18. Geering ADW, Maumus F, Copetti D, Choisne N, Zwicki D, Zytnicki M, McTaggart AR, Scalabrin S, Vezzulli S, Wing RA, Quesneville H, Teycheney PY (2014) Endogenous florendoviral elements are major components of plant genomes and molecular fossils of reverse-transcribing viruses with unique and variable genome organizations. Nat Commun 5:5269
19. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) ClustalW and ClustalX version 2.0. Bioinformatics 23:2947–2948
20. Hasegawa M, Kishino H, Yano T (1985) Dating the human–ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 2:160–174
21. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:870–874