UNIVERSIDADE FEDERAL DE ALAGOAS
REDE NORDESTE DE BIOTECNOLOGIA

GIANCARLO DE BRITO LYRA SANTOS

Diagnóstico e descoberta de badnavírus associados a cará-moela (*Dioscorea bulbifera* L.) no Brasil via sequenciamento Illumina

Maceió, AL
2022

GIANCARLO DE BRITO LYRA SANTOS

Diagnóstico e descoberta de badnavírus associados a cará-moela (*Dioscorea bulbifera* L.) no Brasil via sequenciamento Illumina

Tese apresentada à Universidade Federal de Alagoas como parte das exigências do Programa de Pós-Graduação em Biotecnologia, da Rede Nordeste de Biotecnologia (Renorbio), para obtenção do título de *Doutor em Biotecnologia.*

Maceió, AL

2022

Catalogação na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecária: Taciana Sousa dos Santos – CRB-4 – 2062

S237d Santos, Giancarlo de Brito Lyra.
Diagnóstico e descoberta de badnavírus associados a cará-moela (*Dioscorea bulbifera* L.) / Giancarlo de Brito Lyra Santos. – 2022.
58 f. : il. color.

Orientador: Gaus Silvestre de Andrade Lima.
Coorientadora: Mayra Machado de Medeiros Ferro.
Tese (Doutorado em Biotecnologia) – Universidade Federal de Alagoas. Instituto de Química e Biotecnologia. RENORBIO. Maceió, 2022.

Inclui bibliografias.

1. Cará-moela (*Dioscorea bulbifera Linn*). 2. Plantas alimentícias não convencionais. 3. *Badnavírus*. 4. Filogenia. 5. Sequência genômica. I. Título.

CDU: 633.496

Dedico essa tese à minha família, em especial aos meus pais, Verônica e Juvêncio (*in memoriam*), meus filhos Samuel e Gabriel e minha esposa Rafaela, aos meus amigos, professores e alunos.

AGRADECIMENTOS

Agradeço, primeiramente, ao meu Orientador, Prof. Dr. Gaus Silvestre, por ter aceitado a orientação, ter tido empenho em ajudar e sempre se preocupar com o andamento do projeto. À Prof. Dra. Iraildes Assunção, pela confiança, respeito ao trabalho e disposição de sempre para ajudar.

À coorientadora Dra. Mayra Ferro, braço direito nas realizações de tarefas laboratoriais, com sua experiência e vontade de ajudar a todo momento. Sem ela, tudo seria muito mais difícil e o caminho talvez não fosse o mesmo.

Ao Dr. Frederico Feijó, pelo auxílio nas construções das árvores filogenéticas, desde as aulas até a produção final.

Ao amigo, Prof. Dr. Roberto Ramos, que desde o mestrado esteve sempre presente, atencioso e disposto a ajudar. Suas contribuições na revisão do trabalho, nas análises de resultados e publicações foram primordiais em minha trajetória.

Aos meus pais, Juvêncio Inácio Filho (*in memoriam*) e Verônica Valença de Brito Lira, alicerces maiores de incentivo aos estudos, por mostrarem o caminho que não tinham trilhado e sempre se orgulharem de mim.

À minha esposa Ariana Rafaela (Rafa), pelo incentivo incondicional e por estar ao meu lado em todos os momentos.

Aos meus filhos, Samuel Brito (Samuca) e Gabriel Brito (Biel), por existirem e me apresentarem o amor maior da vida.

Ao meu irmão Beto Brito, meu grande amigo, incentivador, conselheiro e parceiro de todas as horas.

À minha sogra Edisônia, por todo carinho e apoio de sempre.

Aos meus cunhados e cunhadas: Tanis, João, Mônica, Léa, Olivan, Cauper, Alan, Dani e Camila, muito presentes em todos os momentos.

Aos meus familiares e amigos, por compreenderem minha ausência em tantos momentos e mesmo assim continuarem demonstrando amor e fraternidade.

À minha amiga, desde o Colégio Dinâmico (Ensino Médio), Dra. Yamina Coentro, por todo incentivo e ajuda sempre disponível.

Aos meus colegas de laboratório Lívia, Mariote, Luan, Jaque, Nayana, Mayara, Yolanda, Helloá, Lucas, que de alguma forma me orientaram, trocaram ideias, se dispuseram a ajudar, tiveram palavras de incentivo ou até uma boa conversa, sempre torcendo por bons resultados.

Aos que me ajudaram de alguma forma ao longo dessa jornada acadêmica: João Manoel, Adso, Michelle, João Manoel, Ana Paula, e Ailton.

A todos os professores, coordenação e funcionários da Rede Nordeste de Biotecnologia (RENORBIO), por manter acesa a chama da pesquisa e do empreendedorismo.

À Universidade Federal de Alagoas (UFAL), por me abraçar desde 1998, durante a graduação em Odontologia, Mestrado em Ciências e agora como polo de Alagoas do Renorbio, no Doutorado.

Ao Instituto Federal de Alagoas, por me liberar das atividades durante o período de Doutorado, incentivo enorme para qualificação e crescimento de meu trabalho. Irei sempre buscar retribuir à altura com muito trabalho e levando o que há de melhor aos nossos alunos.

Aos membros das bancas de defesa e qualificação, pelas disponibilidades, disposição e contribuições.

Aos meus professores do Ensino Fundamental, Médio, Graduação, Mestrado e Doutorado. Cada um com sua contribuição, de seu jeito, todos fundamentais para minha formação pessoal e acadêmica.

Aos meus alunos do Instituto Federal de Alagoas, em especial aos que tive que deixar durante esse período de qualificação profissional.

A todos os meus alunos, do passado, presente e futuro, que me ensinam que podemos almejar mais e tudo vale a pena quando nos esforçamos, por serem combustível para que eu consiga buscar melhorar a cada dia, por mim e por vocês.

Aos profissionais da saúde que atuaram de forma maestral durante o período de pandemia COVID-19.

Aos profissionais da ciência e pesquisa, por batalharem sempre, mesmo enfrentando diversas adversidades em tempos de desconfiança, falta de apoio, dificuldades financeiras e negacionismo.

A todos os profissionais da educação, guerreiros da batalha diária por uma educação de qualidade, a maioria com salários minúsculos e uma grande força de vontade, buscando o melhor para o aluno.

RESUMO

SANTOS, Giancarlo de Brito Lyra, D.Sc., Universidade Federal de Alagoas, maio de 2022. **Diagnóstico e descoberta via sequenciamento Illumina de badnavírus associados com cará-moela (*Dioscorea bulbifera* L.) no Brasil.** Orientador: Gaus Silvestre de Andrade Lima. Coorientadora: Mayra Machado de Medeiros Ferro.

Dioscorea bulbifera L., popularmente conhecida como cará-moela, é uma cultura pertencente à família Dioscoreaceae e tem despertado interesse devido à sua relevância socioeconômica para pequenos agricultores no Brasil. Apesar de sua resistência natural a doenças, o cará-moela é afetado por pararetrovírus do gênero *Badnavirus* (família *Caulimoviridae*). Os badnavírus possuem genomas de DNA fita dupla, semicirculares, com 7,4 a 9,0 kb, encapsidados em partículas baciliformes não envelopadas e são transmitidos principalmente por cochonilhas (Pseudococcidae). Para caracterizar molecularmente a diversidade de espécies e genomas completos de badnavírus associados a *D. bulbifera* no Brasil, amostras (n=60) foram coletadas em diferentes regiões produtoras. Sequências parciais (~530 pb) dos domínios da transcriptase reversa (RT) e ribonuclease H (RNase H) foram amplificadas via PCR e sequenciadas a partir de 26 amostras positivas, enquanto sequências virais completas foram recuperadas de três amostras representativas. As sequências parciais da RT-RNase H foram identificadas como pertencentes aos badnavírus *Dioscorea bacilliform AL virus*, *Dioscorea bacilliform SN virus* e *Dioscorea bacilliform TR virus*, e ao endógeno de *Dioscorea rotundata* eDBV12. Uma possível espécie nova, provisoriamente denominada *Dioscorea bacilliform BL virus*, foi parcialmente caracterizada, compartilhando maior identidade de nucleotídeos com eDBV12, em 73,4-79,9%. A árvore filogenética baseada em sequências parciais da RT-RNase H mostrou que as novas sequências foram agrupadas em três clados diferentes, com a nova espécie sendo mais próxima de eDBV12. Finalmente, três genomas de badnavírus baseados em sequenciamento *Illumina* foram montados a partir de 2.240 a 30.668 *reads* com cobertura de 74 a 963 vezes. Os novos genomas variaram de 7.208 a 7.420 pb em tamanho e exibiram organização genômica típica de badnavírus com pelo menos três *open reading frames* (ORFs 1-3). As sequências completas da RT-RNase H (~1.230 pb) dos novos genomas virais foram mais intimamente relacionadas ao *Dioscorea bacilliform AL virus* (n=1) e *Dioscorea bacilliform SN virus* (n=2), com 85,1-86,6 e 82,2- 83,9% de identidade nucleotídica, respectivamente. Esses resultados reforçam a alta diversidade de espécies de badnavírus associadas a *Dioscorea* spp. e constituem o primeiro relato de *Dioscorea bacilliform TR virus* (DBTRV) em *D. bulbifera*. Além disso, essas são as primeiras sequências genômicas completas de badnavírus infectando cará-moela no Brasil.

Palavras – chave: Plantas alimentícias não convecionais (PANCs), Fitovirose, Filogenia, *Badnavirus*, *Dioscorea*, *Dioscorea bulbifera*, *Dioscorea bacilliform virus*.

ABSTRACT

SANTOS, Giancarlo de Brito Lyra, D.Sc., Universidade Federal de Alagoas, May of 2022. **Diagnostics and Illumina sequencing discovery of badnaviruses associated with air yam (*Dioscorea bulbifera* L.) in Brazil.** Adviser: Gaus Silvestre de Andrade Lima. Coadviser: Mayra Machado de Medeiros Ferro.

Dioscorea bulbifera L., commonly known as air yam, is an edible crop belonging to the botanical family Dioscoreaceae and it has increasingly attracted attention due to its socioeconomical relevance for smallholder farmers in Brazil. Although its natural resistance to diseases, air yam is affected by plant pararetroviruses into the genus *Badnavirus* (family *Caulimoviridae*). Badnaviruses have double-stranded, semicircular, DNA genomes of 7.4 to 9.0 kb in size, encapsidated into non-enveloped, bacilliform particles and are mainly transmitted by mealybugs (Pseudococcidae). To molecularly characterize the species diversity and complete genomes of badnaviruses associated with *D. bulbifera* in Brazil, plant samples (n=60) were collected from different growing regions. Partial sequences of the reverse transcriptase (RT) and ribonuclease H (RNase H) domains (~530 bp) were amplified by PCR and Sanger sequenced from 26 positive samples, while full-length genome sequences were recovered from three representative samples. The new partial RT-RNase H sequences were identified as belonging to the badnaviruses *Dioscorea bacilliform AL virus*, *Dioscorea bacilliform SN virus*, and *Dioscorea bacilliform TR virus*, and the endogenous pararetrovirus *Dioscorea rotundata* endogenous virus eDBV12. Further, a putative species novel tentatively named *Dioscorea bacilliform BL virus* was partially characterized, sharing highest nucleotide identity with eDBV12 sequences, at 73.4-79.9%. The Bayesian phylogenetic tree based on partial RT-RNase H sequences showed that the new sequences were clustered in three different clades, with the new species being more closely related to eDBV12. Finally, three Illumina-based badnaviral genomes were assembled from 2,240 to 30,668 reads and a coverage depth of 74 to 963 times. The new genomes ranged from 7,208 to 7,420 bp in size and showed typical badnaviral genomic organization with three main open reading frames (ORFs 1-3). The complete RT-RNase H sequences (~1,230 bp) retrieved from the new viral genomes were more closely related to *Dioscorea bacilliform AL virus* (n=1) and *Dioscorea bacilliform SN virus* (n=2), at 85.1-86.6 and 82.2-83.9% nucleotide identity, respectively. These results reinforce the high badnaviral species diversity usually observed associated with *Dioscorea* spp. and constitute the first report of *Dioscorea bacilliform TR virus* (DBTRV) in *D. bulbifera*. Also, these are the first complete genome sequences of yam-infecting badnaviruses described in Brazil.

KEYWORDS: Non-conventional food plants (NCFPs), Phytovirus, Phylogeny, *Dioscorea*, *Dioscorea bacilliform virus*.

SUMÁRIO

INTRODUÇÃO

Espécies cultivadas pertencentes ao gênero *Dioscorea* são de importância socioeconômica em todo o mundo, especialmente para pequenos agricultores. *Dioscorea bulbifera* L., popularmente conhecida como cará-moela, é uma planta trepadeira da família Dioscoreaceae que produz tubérculos comestíveis e bulbilhos aéreos, sendo comumente cultivada em sistemas de agricultura familiar (CROXTON et al., 2011). Essa espécie tem chamado atenção devido à sua classificação na categoria de plantas alimentícias não convencionais, também conhecidas como PANCS (KINUPP; LORENZI, 2014), por sua capacidade de armazenamento por longos períodos e resistência a pragas e doenças, garantindo assim a segurança alimentar (FERREIRA, 2011) e pela ação fitoterápica, com efeito terapêutico contra diversas doenças ao redor do mundo (KUNDU et al., 2021). O centro de diversidade genética de *D. bulbifera* parece ser tanto a Ásia quanto a África e, atualmente, está espalhado pelas Américas (LEBOT, 2019). No Brasil, é considerada uma planta exótica, de fácil reprodução por propagação vegetativa e bem adaptada a diferentes regiões de cultivo (HAMMER, 1998; GOVAERTS et al., 2007; MAURIN et al., 2016).

Os vírus pertencentes ao gênero *Badnavirus* (família *Caulimoviridae*) possuem genoma de DNA de fita dupla, semicircular, de 7,4 a 9,0 kb de tamanho, encapsidados em partículas baciliformes não envelopadas com dimensões de 30 nm x 130 a 150 nm e são considerados pararetrovírus de plantas, que replicam através de um intermediário de RNA (HARPER; HULL, 1998; GEERING; HULL, 2012). Os badnavírus são transmitidos principalmente por espécies de cochonilhas (Hemiptera: Pseudococcidae), de forma semipersistente, sendo detectados em plantas sintomáticas ou assintomáticas (GEERING; HULL, 2012; BHAT et al., 2016).

Partículas semelhantes a badnavírus foram relatadas pela primeira vez em *Dioscorea* spp. na década de 1970 (HARRISON; ROBERTS, 1973; MANTELL; HAQUE, 1979). A caracterização molecular de sequências parciais e completas do genoma viral associado a plantas de *D. alata*, originalmente coletadas na Nigéria, confirmaram a presença de uma nova espécie de badnavírus conhecida como *Dioscorea bacilliform AL virus* (PHILLIPS et al., 1999; BRIDDON et al., 1999). Um segundo badnavírus infectando *Dioscorea sansibarensis* foi descrito posteriormente, conhecido como *Dioscorea bacilliform SN virus* (SEAL; MULLER, 2007). Esses badnavírus estão disseminados nas principais áreas de cultivo em todo o mundo, infectando indiscriminadamente diferentes espécies de *Dioscorea* (ENI et al., 2008; KENYON et al., 2008; BOUSALEM et al., 2009). Pelo menos seis espécies adicionais de badnavírus, associadas a *Dioscorea* spp., são reconhecidos pelo *International Comitee on Taxonomy of*

Viruses (ICTV, talk.ictvonline.org), *Dioscorea bacilliform AL virus 2*, *Dioscorea bacilliform ES virus*, *Dioscorea bacilliform RT virus 1*, *Dioscorea bacilliform RT virus 2*, *Dioscorea bacilliform RT virus 3* e *Dioscorea bacilliform TR virus*. Além disso, plantas de inhame pertencentes ao complexo de espécies *Dioscorea cayenensis-rotundata* que abrigam diferentes grupos de sequências pararetrovirais endógenas (EPRVs) foram relatadas (BOUSALEM et al., 2009; SEAL et al., 2014; UMBER et al., 2014).

Portanto, o presente estudo teve como objetivo diagnosticar e estudar a diversidade de badnavírus infectando cará-moela (*Dioscorea bulbifera*) em diferentes regiões do Brasil, através de amplificação por PCR e sequenciamento dos domínios da transcriptase reversa (RT) e da ribonuclease H (RNase H), bem como sequenciar genomas completos de badnavírus infectando *D. bulbifera* no Nordeste do Brasil, através de sequenciamento de alto rendimento.

CAPÍTULO 1

REVISÃO DE LITERATURA

Cará-moela *(Dioscorea bulbifera)*

O gênero *Dioscorea* possui o maior número de representantes na família Dioscoreaceae e várias espécies são de importância socioeconômica. *Dioscorea* spp. são comuns em regiões tropicais, subtropicais e temperadas (MONTALDO, 1991), com muitas variedades de inhame sendo introduzidas na América do Sul pelos portugueses e espanhóis no século XVI durante a colonização (LEBOT, 2019).

Espécies do gênero *Dioscorea* podem ser classificadas como domesticadas e selvagens, distinguindo-se também como culturas medicinais ou comestíveis (LEBOT, 2019). A maior parte de seu cultivo é de interesse econômico ligado à alimentação, uma vez que o inhame é um alimento básico importante para milhões de pessoas em todo o Mundo (LUO et al., 2022).

Globalmente, o inhame é a quarta cultura de raízes e tubérculos mais importante em produção após batata, batata doce e mandioca (LUO, et al., 2022). De acordo com a FAO (Food and Agriculture Organization of the United Nations), em 2020, a produção mundial de inhame foi de 74,8 milhões de toneladas, sendo o Brasil responsável por cerca de 250 mil toneladas. Segundo Lebot (2019), as espécies de inhame mais cultivadas no mundo são *D. alata*, *D. cayenensis* e *D. rotundata*, enquanto *D. bulbifera*, *D. dumetorum*, *D. esculenta*, *D. japonica*, *D. nummularia*, *D. pentaphylla*, *D. transversa*, *D. oppositifolia* e *D. trifida* são enquadradas como cultivos menores.

D. bulbifera é uma espécie de trepadeira que pode atingir seis metros de altura, produzindo tubérculos e bulbilhos aéreos. A origem de *D. bulbifera* parece ser tanto na Ásia quanto na África, mas é amplamente distribuída em regiões temperadas, África tropical, norte da Austrália, subcontinente indiano, Estados Unidos, China, Camboja, Tailândia, Malásia e Laos (LIU et al., 2010; CASTRO et al., 2012; GUAN et al., 2017). No Brasil, essa espécie é conhecida popularmente como inhame-do-ar, cará-moela (devido à sua semelhança com a moela de galinha), inhame borboleta ou inhame de barbante (SILVA et al. 2020)

Segundo Correa (1978), *D. bulbifera* é uma trepadeira robusta, de caule herbáceo, cilíndrico, estriado, enrolado no sentido anti-horário, com folhas alternadas, de pecíolo longo e lâmina muito desenvolvida, perfeitamente cordiforme, 12-18 cm de comprimento e 10-15 cm de largura. A inflorescência masculina é paniculada, 12 cm ou mais, composta, e a inflorescência feminina está contida em longas espigas axilares de 18-25cm de tamanho. As

flores são sésseis, solitárias, pequenas, pouco visíveis com peritônio violeta e seis estames férteis. O fruto é uma cápsula oblonga, e cada lóculo contém duas sementes aladas na parte inferior. Esta espécie é uma das mais importantes da família e é facilmente distinguida das demais espécies pelos numerosos bulbilhos aéreos, que nascem nas axilas das folhas e variam em forma e tamanho, sendo a reprodução vegetativa seu principal mecanismo de propagação (Figura 1).

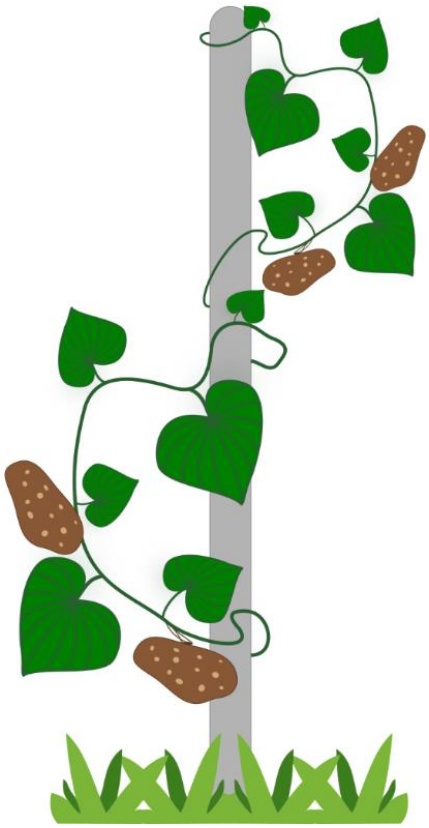


Figura 1. Esquema de uma planta de cará-moela (*Dioscorea bulbifera*) com bulbilhos tipo moela também conhecidos como tubérculos aéreos, formados a partir da base dos pecíolos. Fonte: o autor.

O cará-moela, no Brasil, ainda pode ser considerada uma planta alimentícia não convencional - PANC (KINUPP; LORENZI, 2014). Essa planta tem sido utilizada há muito tempo como medicamento tradicional com efeito terapêutico contra diversas doenças ao redor do mundo (KUNDU et al., 2021). *D. bulbifera* possui uma ampla gama de atividades biológicas devido à sua diversidade química (IKIRIZA et al., 2019), os quais podem variar de acordo com a localização geográfica e a parte da planta, além do tipo de solvente usado no processo de extração (IKIRIZA et al., 2019). Estudos mostraram a presença de vários componentes, como

alcaloides, carboidratos e proteínas, o que indica atividade antioxidante em *D. bulbifera*. (BALASUBRAMANIAN et al., 2012).

Apesar de cultivadas mundialmente, ainda são observadas limitações na produtividade de *Dioscorea* spp. No Nordeste do Brasil, onde mais de 90% da produção do país é responsável pelo cultivo, os principais fatores para a baixa produtividade são manejo inadequado da cultura, uso de túberas de baixa qualidade, não uniformidade de tamanho e maturação, injúrias nos tubérculos que facilitam a contaminação por microrganismos do solo, e alto valor de túberas, que representa cerca de 60% do custo de produção (CAZÉ FILHO, 2002). Além disso, a baixa qualidade e produtividade do inhame têm sido associadas a diferentes problemas fitossanitários (MOURA, 2005).

Pragas e doenças, como bactérias, fungos, insetos, nematoides e vírus, têm impactos negativos diretos no rendimento e na qualidade de produção do inhame. Os vírus estão entre os mais graves, pois são os mais difíceis de controlar, disseminam-se facilmente com material de plantio e foram relatados em todas as regiões de cultivo de inhame do mundo (ITA, et al. 2020; LUO et al., 2022).

Doenças virais em inhame foram identificadas pela primeira vez em Serra Leoa e Porto Rico em 1936 (COOK, 1978). Descrições subsequentes de infecção por vírus ocorreram em 1957 (MIEGE, 1957) e 1961 (ROBERTSON, 1961). Sintomas de mosaico leves a graves foram posteriormente relatados em *D. rotundata* na Nigéria (TERRY, 1976). Os vírus também foram relatados infectando diferentes espécies de inhame em todas as regiões onde as culturas foram estabelecidas (ALEMAN-VERDAGUER et al., 1997; HUGHES, 1997; PHILLIPS et al., 1999; SEAL; MULLER, 2007). As viroses são de grande relevância em culturas propagadas vegetativamente como *Dioscorea* spp., resultando em acúmulo, sobrevivência e disseminação do patógeno (KENYON et al., 2001; SEAL; MULLER, 2007). Plantas infectadas perdem vigor e produzem tubos de baixa qualidade (AMUSA et al., 2003). Os sintomas associados às doenças virais variam de acordo com o grupo de vírus, mas geralmente incluem clorose foliar grave, mosaico, cordão de sapato, clorose internerval, nanismo, distorção foliar, entre outros. Esses sintomas, que afetam principalmente a folhagem, levam à redução da capacidade fotossintética da planta infectada com efeitos deletérios na produção, qualidade dos tubérculos e, em alguns casos, podem causar a morte das plantas afetadas (THOUVENEL; DUMONT, 1988).

Família *Caulimoviridae*

A família *Caulimoviridae* engloba vírus de plantas com genoma de DNA de fita dupla (dsDNA) classificados como pararetrovírus e diferem dos retrovírus com base em seu genoma de DNA e integração irregular no genoma do hospedeiro para replicação (TEMIN, 1985). Os vírus desta família possuem partículas não envelopadas, isométricas (50-52 nm de diâmetro) ou baciliformes (30 nm de diâmetro e 130-150 nm de comprimento) (TEYCHENEY et al., 2020). Os vírions contêm uma única molécula de DNA de fita dupla circular não covalentemente fechada com 7,2 a 9,2 kb de tamanho (FAUQUET et al., 2005; TEYCHENEY et al, 2020). O genoma dos caulimovírus contém de um a oito quadros de leitura aberta (ORFs) e a organização genômica é uma das principais características utilizadas para distinguir os diferentes gêneros desta família, além do tipo de inseto vetor, gama de hospedeiros e relação filogenética. Onze gêneros são atualmente reconhecidos pelo ICTV: *Badnavirus, Caulimovirus, Cavemovirus, Dioscovirus, Petuvirus, Rosadnavirus, Ruflodivirus, Solendovirus, Soymovirus, Tungrovirus* e *Vaccinivirus* (https://talk.ictvonline.org/).

Os membros desta família são amplamente distribuídos geograficamente, com a maioria das espécies dos gêneros *Tungrovirus* e *Badnavirus* localizadas em regiões tropicais e subtropicais, e os demais membros encontrados especialmente em áreas temperadas (TEYCHENEY et al, 2020). Os hospedeiros naturais dos caulimovírus são as Angiospermas das classes Dicotyledonae e Monocotyledonae. Dependendo do gênero, a transmissão natural do vírus pode ocorrer via inseto vetor ou por contato entre plantas hospedeiras, bem como por sementes ou grão de pólen e por propagação vegetativa. A transmissão também pode ser realizada por técnicas como inoculação mecânica e enxertia (FAUQUET et al., 2005).

Caulimovírus são classificados como pararetrovírus, um termo introduzido para definir vírus de plantas que se replicam por transcriptase reversa, onde o genoma é diretamente transcrito em RNA mensageiro, que é sintetizado a partir do genoma original e transcrito de volta ao DNA e difere dos retrovírus com base em seu genoma de DNA e integração passiva no genoma do hospedeiro, não sendo necessário codificar uma enzima integrase para completar seu ciclo de replicação (BHAT et al., 2016).

As sequências virais podem ocorrer dispersas no genoma das plantas, originadas de uma infecção viral anterior que se aderiu à linhagem germinativa. Tais sequências são conhecidas como Sequências Pararetrovirais Endógenas (EPRVs) e são a classe mais abundante de sequências virais endógenas em diferentes espécies vegetais (GAYRAL; ISKRA-CARUANA, 2009). Um número crescente de EPRVs de *Caulimoviridae* tem sido identificado no genoma de muitas espécies de plantas (GEERING; SCHARASCHKIN; TEYCHENEY et al. 2020). Os sintomas causados por esses vírus são variáveis e dependem da espécie viral, hospedeiro e

condições climáticas. Mosaico, clareamento de nervuras, clorose entre nervuras e estrias são os sintomas mais frequentes observados em infecções causadas por diferentes gêneros de *Caulimoviridae* (GEERING; HULL, 2012).

Gênero Badnavirus

A origem do nome *Badnavirus* (i.e., vírus baciliforme de DNA) está relacionada à forma baciliforme de sua partícula viral, sendo o *Commelina yellow mottle virus* (ComYMV) historicamente classificado como a espécie tipo do gênero, que foi descrita pela primeira vez em *Commelina diffusa* L., na Ilha de Guadalupe (MIGLIORI; LASTRA, 1978). Os vírus pertencentes ao gênero *Badnavirus* também são classificados como pararetrovírus de plantas. Portanto, seus genomas são transcritos diretamente em um RNA mensageiro que pode ser retrotranscrito para DNA por uma transcriptase reversa. Os genomas semicirculares de DNA de fita dupla (dsDNA) de aproximadamente 7,0 a 9,2 kb de comprimento são encapsulados em partículas baciliformes não envelopadas com dimensões de 30 nm x 130 a 150 nm (NASCIMENTO, et al. 2020).

O genoma dos badnavírus codifica pelo menos três ORFs, referidas como ORFs 1-3 (Figura 2a) (BOUHIDA et al., 1993; HAGEN et al., 1993; GEERING; HULL, 2012). A proteína ORF1 é associada ao vírion (CHENG et al., 1996), e ORF2 codifica uma proteína de ligação ao ácido nucleico (JACQUOT et al., 1996). A maior é a ORF3, que codifica uma poliproteína compreendendo os domínios do capsídeo viral, da proteína de movimento, a aspartato protease, responsável pela clivagem da poliproteína e os domínios da transcriptase reversa (RT) e da ribonuclease H (RNase H), envolvidos na replicação viral (Figura 2b) (MEDBERRY et al., 1990; HARPER; HULL, 1998; GEERING; HULL, 2012; TEYCHENEY et al, 2020). A demarcação das espécies de *Badnavirus* é baseada na gama de espécies hospedeiras, especificidade do vetor e determinação da sequência nucleotídica da região genômica que inclui os domínios RT-RNase H, localizados na ORF3.O ICTV estabeleceu um critério de ≥80% de nucleotídeos identidade para sequências de RT-RNase H (GEERING; HULL, 2012; TEYCHENEY et al, 2020). O gênero *Badnavirus* é o segundo mais numeroso entre os vírus de plantas com genoma de DNA e é o mais numeroso da família *Caulimoviridae*, com 68 espécies atualmente reconhecidas pelo ICTV (https://talk.ictvonline.org/ictv-reports/ictv_online_report), superado apenas pelo *Begomovirus* (família *Geminiviridae*).

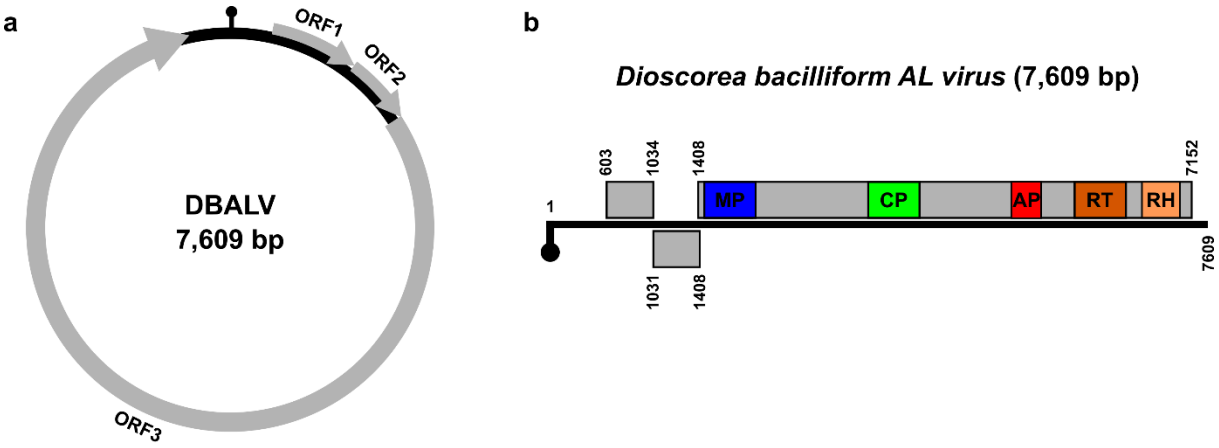


Figura 2. Organização do genoma do *Dioscorea bacilliform AL virus* (DBALV) que infecta inhame. Três quadros de leitura aberta (ORFs 1-3) são codificados pelo genoma viral (a), e a ORF3 compreende importantes domínios conservados, como proteína de movimento (MP), proteína do capsídeo (CP), aspartato protease (AP), transcriptase reversa (RT) e ribonuclease H (RNase H) (Fonte: o autor).

A disseminação dos badnavírus ocorre principalmente por propagação vegetativa em algumas espécies como *Commelina yellow mottle virus, Kalanchoe top-spotting virus, Piper yellow mottle virus, Cocoa swollen shoot virus* e *Taro bacilliform virus*, que também podem ser transmitidos por sementes (HEARON et. al., 1984; DEESHMA et al., 2014; BHAT et al., 2016.). A transmissão natural também pode ser realizada por várias espécies de cochonilhas, como *Sacharicoccus sacchari*, *Planococcus citri*, *Planococcoides njalensis*, *Dysmicoccus neobrevipes*, *Ferrisia virgata* e *Pseudococcus solomonensis*, enquanto as espécies *Gooseberry species vein banding spot virus, Rubus yellow net virus* e *Spiraea yellow leaf virus* são transmitidas por pulgões (BHAT et al., 2016). Alguns desses vírus também podem ser transmitidos experimentalmente por seiva bruta contendo vírus ou por inoculação mediada por *Agrobacterium* (LOCKHART et al., 1996).

Os badnavírus estão amplamente distribuídos em todas as regiões onde o inhame é cultivado (ENI et al., 2008; LUO et al., 2022), e pelo menos oito espécies estão totalmente caracterizadas, *Dioscorea bacilliform AL virus, Dioscorea bacilliform SN virus, Dioscorea bacilliform AL virus 2, Dioscorea bacilliform ES virus, Dioscorea bacilliform RT virus 1, Dioscorea bacilliform RT virus 2, Dioscorea bacilliform RT virus 3* e *Dioscorea bacilliform TR virus* (https://talk.ictvonline.org/ictv-reports). No entanto, estudos realizados na África e no Pacífico Sul sugeriram a ocorrência de uma alta diversidade de badnavírus em inhame com a possível presença de doze espécies diferentes (ENI et al., 2008).

REFERÊNCIAS CITADAS

Aleman-Verdaguer M E G U C, Dubern J, Beachy R N, Fauquet C (1997). Analysis Of The Sequence Diversity Of The P1, Hc, P3, Nib And Cp Genomic Regions Of SeveralYam Mosaic Potyvirus Isolates: Implications For The Intraspecies Molecular Diversity Of Potyviruses. Journal Of General Virology 78: 1253-1264.

Amusa Na, Adegbita A A, Muhammed, S, Daiyewu R. (2003) Yam Diseases And Its Management In Nigeria. African. Journal Of Biotechnology 2: 497-502.

Balasubramanian, J., Dhanalakshmi, R., Jibnomen, J., Manimekalai, P. (2012). A preclinical evaluation on antioxidant and gastroprotective effect of *Dioscorea bulbifera* in Wistar rats. Indian J. Innovations Dev., Vol. 1, No. 3. 149-154.

Bousalem M., Durand O., Scarcelli N. (2009) Dilemmas Caused By Endogenous Pararetroviruses Regarding The Taxonomy And Diagnosis Of Yam (*Dioscorea* Spp.) Badnaviruses: Analyses To Support Safe Germplasm Movement. Arch Virol 154:297–31

Bhat A.I, Hohn T, Selvarajan R. Badnaviruses (2016).: The Current Global Scenario. Viruses, 2016.

Bouhida, M. L.; Lockhart, B. E.; Olszewski, N. E. (1993). An Analysis Of The Complete Sequence Of A Sugarcane Bacilliform Virus Genome Infectious To Banana And Rice. Journal Of E General Virology, V. 74, P.15-22.

Castro, A. P. D., Fraxe, T. D. J. P., Pereira, H. D. S., & Kinupp, V. F. (2012). Etnobotânica Das Variedades Locais Do Cará (*Dioscorea* Spp.) Cultivados Em Comunidades No Município De Caapiranga, Estado Do Amazonas. Acta Botanica Brasilica, V.26, N.3, P.658-667.

Cazé Filho, J. (2002). Clonagem Do Inhame (*Dioscorea* Sp.) Por Métodos Biotecnológicos.In: Simpósio Nacional Sobre As Culturas Do Inhame E Do Taro. João Pessoa, Pb. Anais, Pb: Emepa – Pb, V.1, P. 113-123.

Cheng, C.P.; Lockhart, B.E.; Olszewski, N.E. (1996). The Orf I And Ii Proteins Of Commelina Yellow Mottle Virus Are Virion-Associated. Virology, V. 223, P. 263-271.

Cook A A (1978) Diseases Of Tropical And Sub Tropical Vegetables And Other Plants. New York:Hafner Press, 381p.

Correa M P (1978). Dicionário De Plantas Úteis Do Brasil E Das Exóticas Cultivadas. Rio de Janeiro: Instituto Brasileiro De Desenvolvimento Florestal.

Croxton, M.D., Andreu, M.A., Williams, D.A., Overholt, W.A., Smith, J.A. (2011). Geographic Origins and Genetic Diversity of Air-Potato (Dioscorea bulbifera) in Florida. Invasive Plant Science and Management, 4:22-30.

Deeshma K P, Bhat A.I (2014) Further Evidence Of True Seed Transmission Of Piper Yellow Mottle Virus In Black Pepper (Piper Nigrum L.) Journal Of Plantation Crops. 42 : 289-293.

Eni AO, Hughes JD'A, Rey MEC (2008) Survey of the incidence and distribution of five viruses infecting yams in the major yam-producing zones in Benin. Annals of Applied Biology 153:223-232.

Fauquet C (2005). Virus Taxonomy. Eight Report of The International Committee On Taxonomy Of Viruses. Amsterdam. Elsevier.

Ferreira A B (2011) Sistemas De Cultivo Do Cará *Dioscorea* Spp. Por Sistemas De Cultivo Do Cará *Dioscorea* Spp. Botucatu.

Gayral, P; Iska-Caruana, M. (2009). Phylogeny of Banana streak virus revels recent and repetitive endogenization in the genome of its host (*Musa* sp.). Journal of Molecular Evolution, v. 69. N.1. 65-80.

Geering A D, Scharaschkin T, Teycheney P Y (2020). The Classification And Nomenclature Of Endogenous Viruses Of The Family Caulimoviridae. Archives Of Virology 155: 123-31.

Geering A D W, Hull, R., (2012). Family Caulimoviridae. In: King A M Q, Adams M J, Carstens E B, Lefkowitz E J. Virus Taxonomy. 9th Report Of The InternationalCommittee On Taxonomy Of Viruses. London Uk. Elsevier Academic Press. 429-443.

Guan X R, Zhu L, Xiao Z G, Zhang Y L, Chen H B, Yi T (2017). Bioactivity, Toxicity and Detoxification Assessment of *Dioscorea Bulbifera* L.: A Comprehensive Review. Phytochemistry Reviews, 16: 1-29.

Harper G, Hull R (1998) Cloning And Sequence Analysis Of *Banana Streak Virus* DNA. Virus Genes, 17: 271-278.

Hagen L S, Jacquemond M, Lepingle A, Lot H, Tepfer M (1993) Nucleotide Sequence And Genomic Organisation Of Cacao Swollen Shoot Virus. Virology, 196: 619- 628.

Harrison, B.; Roberts, I. (1973). Association of virus-like particles with internal brown spot of yam (*Dioscorea alata*). Tropical Agriculture 50:335-340.

Hearon S S, Locke J C (1984). . Graft, Pollen And Seed Transmission Of An Agent Associated With Top Spotting In Kalanchoe Blossfeldiana. Phytopathology, 74: 347-50.

Hughes J D A, Dongo L, Atiri G I (1997) Viruses Infecting Cultivated Yams (*Dioscorea alata* and *D. rotundata*) In Nigeria. Phytopathology, 87: 45.

Ikiriza H., Ogwang, P. E., Peter, E. L., Hedmon, O., Tolo, C. U., Abubaker, M., Abdalla, A. A. M. (2019). *Dioscorea bulbifera*, a highly threatened African medicinal plant, a review. Cogent Biology, 5: 1631561. 1-6.

Ita, E.E.; Uyoh, E.A.; Nakamura, I.; Ntui, V.O. (2020). Efficient elimination of *Yam mosaic virus* (YMV) from white yam (*Dioscorea rotundata* Poir.) by cryotherapy of axillary buds. South Afr. J. Bot., 130, 123–129

Jacquot E (1996). The Open Reading Frame 2 Product Of *Cacao swollen shoot* BadnavirusIs A

Nucleic Acid-Binding Protein. Virology, 225: 191-195.

Kenyon L (2008) Yams (*Dioscorea* spp.) From The South Pacific Islands Contain Many Novel Badnaviruses: Implications For International Movement Of Yam Germplasm, Archives Of Virology, 3: 877–889.

Kinupp, V. F., Lorenzi, H (2014). Plantas Alimentícias Não Convencionais (PANC) No Brasil: Guia de Identificação, Aspectos Nutricionais e Receitas Ilustradas. São Paulo: Instituto Plantarum de Estudos da Flora.

Kundu, B. B., Vanni, K., Farheen A., Jha, P., Pandey, D. K., Kumar, V. (2021). *Dioscorea bulbifera* L. (*Dioscoreaceae*): A review of its ethnobotany, pharmacology and conservation needs. South African Journal of Botany. Vol.140. 365-374

Lebot, V. (2019) Tropical root and tuber crops, 2nd edn. Cabi, Wallingford, Oxfordshire, UK ; Boston, MA

Liu W, Wang H, Pang X, Yao W, Gao X (2010). Characterization And Antioxidant Activity Of Two Low-Molecular-Weight Polysaccharides Purified From The Fruiting Bodies Of Ganoderma Lucidum. International Journal Of Biological Macromolecules, 46: 451–457.

Lockhart B E L, Olszewski N E (1996). Serological And Genomic Heterogeneity Of Banana Streak Badnavirus: Implications For Virus Detection In*Musa* germplasm. In: Ganry, J. (Ed.). Breeding Banana And Plantain For Resistance To Diseases And Pests. Montpellier: Cirad/Inibap, 105-113.

Luo, G., Podolyan, A., Kidanemariam, D. B., Pilotti, C., Houliston, G., Sukal, A. C. (2022). A Review of Viruses Infecting Yam (*Dioscorea* spp.). Viruses, 14:662. 1-20.

Mantell, S.H.; Haque, S.Q (1979). Internal brown spot disease of yams. Yam Virus Proj. Bull. (Trinidad Tobago), ill, 1–13.

Medberry S L, Lockhart B E L, Olszewski N E (1990) Properties Of *Commelina yellow mottle virus* Complete Dna Sequence, Genomic Discontinuities And Transcript Suggest That It Is A Pararetrovirus. Nucleic Acids Research, 18: 5505-5513.

Miegi, J. (1957). Influence de Quelques caracteres des tubercules semences sur la levée et le rendement des ignames cultivées. Journal d'Agriculture Tropicale et de Botanique Appliquée, v.4, 315-341.

Migliori, A.; Lastra, R. (1978). Etude de virus present chez *Commelina diffusa* Burm. En Guadeloupe. Annals of Phytopatology, v. 10. 467-477.

Moura R M (2005). Doenças do Inhame-Da-Costa. In: Kimati H, Amorim L, Bergamin Filho A, Camargo L E A, Rezende, J A M. (Eds.). Manual De Fitopatologia. São Paulo: Ceres,. V.2, Doenças Das Plantas Cultivadas. 4ª Ed. Cap. 47: 415-419.

Montaldo A (1991). Cultivo De Raíces Y Tubérculos Tropicales. São José: Instituto

Interamericano De Ciências Agrícolas De La Oea, 408 P.

Nascimento J. P., Oliveira M. L., Peixinho G S, Ferro M. M., Silva S J C, Versey E A, Lima G S A, Assuncao I P (2020). Incidência e Caracterização Molecular de Badnavírus Em Bancos de Germoplasma de Inhame no Brasil. Summa. Phytopathologica, 46: 242-249.

Odu B O (2004). Responses Of White Yam (*Dioscorea Rotundata*) Cultivars To Inoculation With Three Viruses. Plant Pathology, 53: 141–147.

Phillips S (1999). The Partial Characterization of a Badnavirus Infecting The Greater AsiaticOr Water Yam (*Dioscorea alata*). Journal of Phytopathology, 147: 265-269.

Robertson, D G (1961). Notes in annual reports. Federal departamento f Agricultural Research. Nigéria. 1959-1960.

Seal, S., Muller, E. (2007). Molecular analysis of a full-length sequence of a new yam badnavirus from *Dioscorea sansibarensis*. Arch. Virol, 152:819-825.

Seal, S., Turaki, A., Muller, E., Kumar, P.L., Kenyon, L., Filloux, D., Galzi, S., Lopez-montes, A., Iskra-caruana, M. (2014). The prevalence of badnaviruses inWest African yams (*Dioscorea cayenensis-rotundata*) and evidence of endogenous pararetrovirus sequences in thier genomes. Virus Res. 186:144-154.

Silva, E. N. L., Araújo, J. F. da S., Pereira, A. S. , Santos, V. F., Costa, D. M., Pires, C. R. F. (2020). Caracterização nutricional das espécies cará-moela (*Dioscorea Bulbífera* L.) e cará (*Dioscorea* spp.) Revista Desafios –v. 7, n. 3. 357-366.

Temin, H. M. (1985) Reverse transcription in the eukaryotic genome: Retroviruses, pararetroviruses, retrotransposons and retrotranscripts. Molecular Biology and Evolution. V.2. 455-468.

Terry, E. R. (1976). Incidence, symptomatology and transmission of a yam virus in Nigeria. In: proceedings of the 4th Symposium of the International Society for Tropical Root Crops. Cali. Ed. J. Cook, R. Mcintyre e M. Graham. 170-173.

Teycheney PY, Geering AD, Dasgupta I et al (2020) ICTV Virus taxonomy profile: *Caulimoviridae*. J Gen Virol 101(10):1025

Thouvenel, J.C., Dumont, R. (1988). An Epidemiological Approach to The Study of Yam Mosaic Viruses In The Ivory Coast. Proc. Int. Soc. Tropical Root Crops. 45:643-649.

Umber, M., Filloux, D., Muller, E., Laboureau, N., Galzi, S., Roumagnac, P., Pavis, C., Teycheney, P., Seal, S.E. (2014). The genome of African yam (Dioscorea cayenensis-rotundata complex) hosts endogenous sequences from four distinct badnavirus species. Mol. Plant Pathol. 15:790-801.

CAPÍTULO 2

BADNAVIRUS SEQUENCE DIVERSITY REVEALS ONE PREVIOUSLY UNCHARACTERIZED BADNAVIRUS ASSOCIATED WITH AIR YAM (*Dioscorea bulbifera* L.) IN BRAZIL

Giancarlo B.L. Santos, Mayra M.M. Ferro, Frederico M. Feijó, Roberto Ramos-Sobrinho, Iraildes P. Assunção, Gaus S.A. Lima. Badnavirus sequence diversity reveals one previously uncharacterized badnavirus associated with air yam (*Dioscorea bulbifera* L.) in Brazil. Tropical Plant Pathology, *submitted*.

Badnavirus sequence diversity reveals one previously uncharacterized badnavirus associated with air yam (*Dioscorea bulbifera* L.) in Brazil

Giancarlo B.L. Santos[1,2], Mayra M.M. Ferro[2], Frederico M. Feijó[2], Roberto Ramos-Sobrinho[2*], Iraildes P. Assunção[2], Gaus S.A. Lima[2*]

[1]Campus Coruripe, Instituto Federal de Alagoas, 57230-000, Coruripe, Alagoas, Brazil
[2]Campus de Engenharias e Ciências Agrárias, Universidade Federal de Alagoas, 57100-000, Rio Largo, Alagoas, Brazil

***Corresponding authors:** R.R.S. (robertorsobrinho@gmail.com) and G.S.A.L. (gaus@ceca.ufal.br)

Abstract

Dioscorea bulbifera L., commonly known as air yam, is an edible crop belonging to the botanical family Dioscoreaceae, and it has increasingly attracted attention due to its socioeconomical relevance for smallholder farmers in Brazil. Although its natural resistance to diseases, air yam is affected by plant pararetroviruses into the genus *Badnavirus* (family *Caulimoviridae*). To molecularly characterize the species diversity of badnaviruses associated with *D. bulbifera* in Brazil, plant samples (n=60) were collected from different growing regions. Partial sequences of the reverse transcriptase (RT) and ribonuclease H (RNase H) domains were amplified by PCR and Sanger sequenced from 26 positive samples. The new sequences were identified as belonging to the badnaviruses *Dioscorea bacilliform AL virus*, *Dioscorea bacilliform SN virus*, and *Dioscorea bacilliform TR virus*, and the endogenous pararetrovirus dioscorea rotundata endogenous virus eDBV12. Further, a putative species novel tentatively named *Dioscorea bacilliform BL virus* was partially characterized. The Bayesian phylogenetic tree based on partial RT-RNase H sequences showed that the new sequences were clustered in three different clades, with the new species being more closely related to eDBV12. These results reinforce the high badnaviral species diversity usually observed associated with *Dioscorea* spp. and constitute the first report of dioscorea bacilliform TR virus (DBTRV) in *D. bulbifera*.

Keywords: *Dioscorea*, dioscorea bacilliform viruses, *Caulimoviridae*, EPRVs.

1. Introduction

Dioscorea bulbifera L. is an herbaceous twining vine into the botanical family Dioscoreaceae and is commonly called as air yam. The center of genetic diversity of *D. bulbifera* seems to be both Asia and Africa, and it is currently spread throughout the Americas (Hammer, 1998; Govaerts et al., 2007; Maurin et al., 2016). Air yam is a climbing plant, which produces edible underground tubers and aerial bulbils, that can be easily reproduced by vegetative propagation (Croxton et al., 2011). *Dioscorea bulbifera* is also classified into the category of non-conventional edible plants (Kinupp and Lorenzi, 2014). In Brazil, crop plants belonging to the genus *Dioscorea* are socioeconomically important, being mainly cultivated by smallholder farmers, and are considered an alternative crop for food security (Ferreira, 2011).

Viruses belonging to the genus *Badnavirus* (family *Caulimoviridae*) have semicircular, double-stranded DNA (dsDNA) genomes of 7.0-9.2 kbp in size, encapsidated in a non-enveloped bacilliform particle, and encode three to seven open reading frames (ORFs). Badnaviruses replicate through an RNA intermediate molecule, being referred to as plant pararetroviruses, and are mainly transmitted by mealybug species (Pseudococcidae) or, in some instances, aphids, in a non-circulative, semipersistent manner (Geering and Hull, 2012; Bhat et al., 2016).

Badnavirus-like particles were initially described in *Dioscorea* spp. in the 1970s (Harrison and Roberts, 1973; Mantell and Haque, 1979). Thereafter, partial and complete genome sequences obtained from *Dioscorea alata* plants originally collected in Nigeria showed that it represented a new badnaviral species currently known as *Dioscorea bacilliform AL virus* (Phillips et al., 1999; Briddon et al., 1999). A second badnavirus associated with *Dioscorea sansibarensis* was later characterized, at the genome level, and is presently known as *Dioscorea bacilliform SN virus* (Seal and Muller, 2007). These badnaviruses are widespread in the main yam cultivation areas worldwide, indiscriminately infecting different *Dioscorea* species (Eni et al., 2008; Kenyon et al., 2008; Bousalem et al., 2009). At least six other badnaviral species associated with *Dioscorea* spp. are recognized by the International Committee on Taxonomy of Viruses (ICTV), *Dioscorea bacilliform AL virus 2*, *Dioscorea bacilliform ES virus*, *Dioscorea bacilliform RT virus 1*, *Dioscorea bacilliform RT virus 2*, *Dioscorea bacilliform RT virus 3*, and *Dioscorea bacilliform TR virus* (https://talk.ictvonline.org/ictv-reports). Also, yam plants belonging to the *Dioscorea cayenensis-rotundata* species complex harboring different groups of endogenous pararetroviral sequences (EPRVs) have been reported (Bousalem et al., 2009; Seal et al., 2014; Umber et al., 2014).

Here, the species diversity of badnaviruses infecting *D. bulbifera* from different growing regions in Brazil was assessed by PCR amplification and sequencing of the reverse transcriptase (RT) and ribonuclease H (RNase H) domains. At least three previously reported badnaviruses were found, *Dioscorea bacilliform AL virus* (DBALV), *Dioscorea bacilliform SN virus* (DBSNV), and *Dioscorea bacilliform TR virus* (DBTRV). Sequences belonging to a putative new badnaviral species were recovered, and tentatively named as *Dioscorea bacilliform BL virus*. Further, badnavirus-like endogenous sequences were characterized, being most closely related to *Dioscorea rotundata* endogenous pararetrovirus. Together, these results provide important clues on badnaviral diversification in *D. bulbifera* in Brazil, which can be used for development of reliable diagnostic tools, and by breeding programs searching for genetically resistant materials.

2. Materials and Methods

Plant samples

Bulbils of air yam (*D. bulbifera*) were collected from different geographical regions in Brazil. These materials were planted at the experimental field of the Federal University of Alagoas, Rio Largo, Alagoas state, Brazil. Then, leaf samples were collected from each symptomatic and asymptomatic plants and kept at -80^{o}C until being analyzed.

Badnavirus detection

Total DNA was individually extracted from 100-200 mg of frozen leaf tissue using the method described by Doyle & Doyle (1987) and used as template for PCR amplification. The degenerated primers BadnaFP (5'-ATGCCITTYGGIITIAARAAYGCICC-3') and BadnaRP (5'-CCAYTTRCAIACISCICCCCAICC-3'), which amplify the RT-RNase H domains of *Badnavirus* species (Yang et al., 2003), were used for virus detection. The PCR reactions were performed in a total volume of 15 µL, containing 1.5 µL of 10x buffer (100 mM KCl, 100 mM Tris-HCl pH 9.0, 1% Triton-X), 1.2 µL of 2.5 mM dNTPs, 0.4 µL 50 mM MgCl2, 0.2 µL of Taq DNA polymerase, 1.0 µL of each oligonucleotide (10 µM), 1.0 µL (50ng) of total DNA, and 8.7 µL of nuclease-free water. The amplification conditions were: initial denaturation at 94°C for 4 minutes, 35 cycles of denaturation at 94°C for 30 seconds, annealing at 50°C for 30 seconds, and extension at 72°C for 1 minute, and a final extension step at 72°C for 10 minutes. The PCR amplification products were analyzed in 1% agarose gel, stained with ethidium bromide and visualized under ultraviolet light. Expected size amplicons (~580 bp) were purified using the GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare, Illinois, USA)

according to the manufacturer´s protocol and sent for Sanger sequencing at Macrogen Inc. (Seoul, South Korea).

Sequence analysis

The contigs corresponding to the RT-RNase H nucleotide sequences were assembled and edited using CodonCode Aligner v. 4.1.1 (www.codoncode.com). The consensus sequences were initially analyzed with the BLASTn algorithm (Altschul et al., 1990) to identify their closest matches among virus sequences available in the NCBI non-redundant GenBank database (https://www.ncbi.nlm.nih.gov/genbank). Then, similar sequences obtained from GenBank (Supplementary Table S1) were used for species demarcation of the new isolates via pairwise nucleotide sequence comparisons using Sequence Demarcation Tool v. 1.2 (Muhire et al., 2014). The *Badnavirus* species demarcation criterion of ≥80% nucleotide identity for the RT-RNase H domains established by the ICTV was adopted (Geering and Hull, 2012).

Phylogenetic analysis

The RT-RNase H nucleotide sequences of the new isolates and badnaviral sequences retrieved from GenBank were aligned using the MUSCLE algorithm (Edgar, 2004), and manually adjusted in MEGA 7.0 (Kumar et al., 2015). The phylogenetic relationship was determined by Bayesian Inference (BI) through the CIPRES web portal (Miller et al., 2010) using MrBayes v.3.2.3 (Ronquist et al., 2012), assuming a general time reversible (GTR) nucleotide substitution model with a gamma (G) model of rate heterogeneity and invariable (I) sites, determined using MrModeltest 2.3 (Posada and Buckley, 2004) according to the Akaike Information Criterion (AIC). The analysis consisted of two replicates with four chains each for 10 million generations and sampling every 1,000 generations. The first 2,500 trees per run were discarded as a burn-in. The posterior probability values (Rannala and Yang, 1996) were determined from the majority-rule consensus tree reconstructed with the 15,000 remaining trees. The BI tree was edited in FigTree v.1.4 (ztree.bio.ed.ac.uk/software/figtree) and Inkscape (https://inkscape.org/pt/).

3. Results and Discussion

A total of 60 air yam bulbils were obtained from different locations in Brazil, which comprise the gene bank collection of *D. bulbifera* established at the Campus of Engineering and Agrarian Sciences of the Federal University of Alagoas. Leaf samples were collected from symptomatic and asymptomatic plants generated from the air yam bulbils and individually tested, by PCR, using the BadnaFP/BadnaRP primer pair (Yang et al., 2003). Expected size

amplicons of approximately 580 bp were observed from 39 of 60 plants, suggesting an incidence level of 65% (Supplementary Table S2). High incidence of badnaviruses, ranging from 72.0-93.3%, affecting *Dioscorea* spp. have been previously observed in Brazil (Lima et al., 2013; Guimaraes et al., 2015; Nascimento et al., 2020), demonstrating that badnaviruses are widespread in commercial plantations and germplasm collections of yams. Although the PCR primers described by Yang et al. (2003) are well known for being unable to distinguish between episomal and integrated RT-RNase H sequences, it is an important and frequently used tool for badnavirus detection (Geering and Hull, 2012; Bhat et al., 2016; Luo et al., 2022). To avoid spurious diagnostics results, the PCR amplification products were bidirectionally Sanger sequenced and showed that at least 12 of 26 plants were infected by badnavirus species largely known for their episomal forms.

The isolates DBJ2 and DBJ3 shared 99.0% nucleotide identity with each other, and showed highest identity with DBALV sequences, at 84.3-95.2%, while the isolates DBMT1 and DBMT2 showed 99.8% identity between them and were more closely related to DBTRV, at 86.0-86.2% identity. The isolates DBAM1, DBJG, DBT6, DBC2, DBM6, DBM8, DBM9, and DBMU showed 97.9-99.4% nucleotide identity among them, and shared greater identity with DBSNV, at 82.3-83.2%. The isolates DB11, DBM123, DB31, DB32, DBB2, DBCO4, DBE, DBCU2, and DBH showed 82.6-99.1% nucleotide identity with one another, and shared highest nucleotide identity with the *Dioscorea rotundata* endogenous pararetrovirus eDBV12, at 75.7-79.9% identity, suggesting the new sequences may represent a putative new badnaviral species for which the name *Dioscorea bacilliform BL virus* is tentatively proposed (Table 1). Additional studies are needed to determine if this new species represents an episomal or integrated badnavirus. Finally, the isolates DB21, DBB1, DBG, DBT2, and DBT3 shared 99.4-100.0% with one another, and were identified as eDBV12, at 88.6-89.2% nucleotide identity (Table 1). The new sequences reported here were deposited in NCBI-GenBank under accession nos. OM628722-OM628747 (Table 2).

Badnaviruses are present in tropical and subtropical crops, including *Dioscorea* spp., of great socioeconomical importance worldwide and can lead to economic losses between 10-90% (Phillips et al., 1999; Briddon et al., 1999; Seal and Muller, 2007; James et al., 2011; Eni et al., 2008; Kenyon et al., 2008; Bousalem et al., 2009; Silva et al., 2015; Deeshma and Bhat, 2015; Bhat et al., 2016; Luo et al., 2022). Yam plants affected by badnaviruses usually exhibit disease symptoms such as leaf chlorosis and distortion, and dwarfism, which can lead to a reduction in the photosynthetic capacity of the infected plant with deleterious effects on production, tuber quality, and plant death (Thouvenel and Dumont, 1988; 1990). Recently, a taxonomic positioning study in *Badnavirus* suggested partial RT-RNase H sequences (~579 bp) are

sufficient for species demarcation (Ferreira et al., 2019). Therefore, based on the ICTV-approved ≥80% nucleotide identity species demarcation criterion for RT-RNase sequences into the genus *Badnavirus* (Geering and Hull, 2012), isolates of DBALV, DBSNV, DBTRV, and dioscorea bacilliform BL virus (DBBLV) were found to be largely spread in *D. bulbifera* growing-areas in Brazil, reinforcing this host harbors a high badnaviral species diversity that can negatively impact the disease management. To our knowledge, this is the first report of DBTRV in *D. bulbifera* worldwide.

Endogenous pararetroviral sequences (EPRVs) have been shown to be integrated into the genome of the African yam, *D. cayenensis-rotundata* complex, but no evidence of EPRVs has been found in other yam species such as *D. alata* and *D. sansibarensis* (Bousalem et al., 2009; Seal et al., 2014; Umber et al., 2014). In the present study, EPRV sequences (eDBV12) previously reported in *D. cayenensis-rotundata* were also characterized from *D. bulbifera* samples using PCR primers amplifying the badnaviral RT-RNase H domains. These results emphasize the importance of EPRVs present in the genome of yams for implementation of reliable molecular detection tools and, although no evidence of infectious EPRVs has been found in *Dioscorea* spp., it may represent a challenge for yam germplasm conservation and exchange of genetic materials between breeding programs.

The BI phylogenetic tree based on partial RT-RNase H sequences showed that the new sequences were clustered in three different clades (Figure 2). The isolates reported in the present study that shared highest nucleotide identity with *Dioscorea* endogenous sequences were clustered into two subgroups, referred to as subclades 1a and 1b. The subclade 1a is comprised by isolates representing the putative new species DBBLV (isolates DB11, DBM123, DB31, DB32, DBB2, DBCO4, DBE, DBCU2, and DBH), while the isolates DB21, DBB1, DBG, DBT2, and DBT3 clustered in a monophyletic group with eDBV12 sequences, forming a sister clade with DBBLV (subclade 1b; Figure 2). These results agree with the pairwise sequence comparisons and reinforce the sequences in subclade 1a may represent a new badnaviral species. However, since only partial sequences were obtained here, and these isolates were more closely related to eDBV12 endogenous sequences, additional studies are necessary to clarify their episomal or integrated origin.

The isolates DBJ2 and DBJ3 grouped together with other DBALV sequences in the phylogenetic subclade 1c, while the new DBSNV (isolates DBAM1, DBJG, DBT6, DBC2, DBM6, DBM8, DBM9, and DBMU) and DBTRV (isolates DBMT1 and DBMT2) sequences were placed in divergent phylogenetic groups, clades 2 and 4, respectively (Figure 2). So far, eight distinct *Dioscorea*-infecting badnaviruses have been characterized at the genome level (Briddon et al., 1999; Seal and Muller, 2007; Bomer et al., 2016; Umber et al., 2016; Sukal et

al., 2017; Bomer et al., 2018; Sukal et al., 2020), with DBALV and DBSNV being previously reported associated with *D. bulbifera* (Sukal et al., 2020; Nascimento et al., 2020). Although high badnaviral species diversity associated with *D. bulbifera* was observed here, additional samples from different growing regions must be analyzed, as well as the complete genomes need to be fully characterized. Also, assessing the extant species diversity of badnaviruses infecting *D. bulbifera*, and other *Dioscorea* species, and estimating the evolutionary mechanisms acting on diversification of these viruses are important steps to improve disease identification and management.

References

Altschul, S.F. Basic Local Alignment Search Tool. J. of Mol. Bio., 1990, 215:403-410.

Bhat, A., Hohn, T., Selvarajan, R. Badnaviruses: The current global scenario. Viruses, 2016, 8:177.

Bömer, M., Rathnayake, A.I., Visendi, P., Silva, G., Seal, S.E. 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. Arch. Virol. 2018, 163:533-538.

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

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

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

Croxton, M.D., Andreu, M.A., Williams, D.A., Overholt, W.A., Smith, J.A. Geographic Origins and Genetic Diversity of Air-Potato (*Dioscorea bulbifera*) in Florida. Invasive Plant Science and Management, 2011, 4:22-30.

Deeshma, K.P., Bhat, A.I. Complete Genome Sequencing of Piper Yellow Mottle Virus Infecting Black Pepper, Betelvine, And Indian Long Pepper. Virus Genes, 2015, 50:172-175.

Doyle, J.J., Doyle, J.L. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochem. Bull., 1987, 19:11-15.

Edgar, R.C. MUSCLE: a multiple sequence alignment method with reduced time and space

complexity. BMC Bioinformatics, 2004, 5:1-19.

Eni, A.O., Hughes, J.D.A., Asiedu, R., Rey, M.E.C. Sequence diversity among badnavirus isolates infecting yam (*Dioscorea* spp.) in Ghana, Togo, Benin and Nigeria. Arch. Virol. 2008, 153:2263-2272.

Ferreira, A.B. Sistemas De Cultivo Do Cará *Dioscorea* Spp. Por Sistemas De Cultivo Do Cará *Dioscorea* Spp. Botucatu, 2011.

Ferreira, C.H.L.H, Jordão, L.J., Ramos-Sobrinho, R., Ferro, M.M.M., Silva, S.J.C., Assunção, I.P., Lima, G.S.A. Diversification into the genus *Badnavirus*: phylogeny and population genetic variability. Ciência Agrícola, 2019, 17: 59-72.

Geering, A.D.W, Hull, R. Family Caulimoviridae. In: King A.M.Q., Adams M.J., Carstens E.B., Lefkowitz, E.J., (2012) Virus Taxonomy. 9th Report of The InternationalCommittee On Taxonomy Of Viruses. London Uk. Elsevier Academic Press., 2012, 429-443.

Govaerts, R., Wilkin, P., Saunders, R.M.K. World checklist of Dioscoreales. Yams and their allies. The Board of Trustees of the Royal Botanic Gardens, Kew, London, 2007, 1-65.

Guimarães, K.M., Silva, S.J.C., Melo, A.M., Ramos-Sobrinho, R., Lima, J.S., Zerbini, F.M., Assunção, I.P., Lima, G.S.A. Genetic variability of badnaviruses infecting yam (*Dioscorea* spp.) in northeastern Brazil. Tropical Plant Pathology, 2015, 40:111-118.

Hammer, R.L. Diagnosis: *Dioscorea*. Wildland Weeds, 1998, 2:8-10.

Harrison, B., Roberts, I. Association of virus-like particles with internal brown spot of yam (*Dioscorea alata*). Trop. Agric. 1979, 50:335-340.

James, A.P., Geijskes, R.J., Dale, J.L., Harding, R.M. Development of A Novel Rolling-Circle Amplification Technique To Detect Banana Streak Virus That Also Discriminates Between Integrated And Episomal Virus Sequences. Plant Disease, 2011, 95:57-62.

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

Kinupp, V.F., Lorenzi, H. Plantas alimentícias não convencionais (PANC) no Brasil: guia de identificação, aspectos nutricionais e receitas ilustradas. Plantarum, Nova Odessa, 2014, 768.

Kumar, S., Stecher, G., Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7. Mol. Biol. Evol., 2015, 33:1870-1874.

Lima, J.S., Lima, A.T.M., Castillo-Urquiza, G.P., Silva, S.J.C., Assunção, I.P., Michereff, S.J., Zerbini, F.M., Lima, G.S.A. Variabilidade genética de isolados de badnavírus infectando inhame (*Dioscorea* spp.) no Nordeste do Brasil. Tropical Plant Pathology, 2013, 38:349-353.

Luo, G.F., Podolyan, A., Kidanemariam, D.B., Pilotti, C., Houliston, G., Sukal, A.C. A Review of Viruses Infecting Yam (*Dioscorea* spp.). Viruses, 2022, 14:662.

Mantell, S.H., Haque, S.Q. Internal brown spot disease of yams. Yam Virus Proj. Bull. (Trinidad Tobago) 1979, ill, 1–13.

Maurin, O., Muasya, A.M., Catalan, P., Shongwe, E.Z., Viruel, J., Wilkin, P., van der Bank, M. Diversification into novel habitats in the Africa clade of *Dioscorea* (Dioscoreaceae): erect habit and elephant's foot tubers. BMC Evol. Biol., 2016, 16:238.

Miller, M.A., Pfeiffer, W., Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Gateway Computing Environments Workshop (GCE), New Orleans, LA, 2010. Ieee, 1-8.

Muhire, B.M., Varsani, A., Martin, D.P. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. PLoS ONE, 2014, 9:e108277.

Nascimento, J.P., Oliveira, M.L., Peixinho, G.S., Ferro, M.M.M., Silva, S.J.C., Versey, E.A., Lima, G.S.A., Assunção, I.P. Incidência E Caracterização Molecular De Badnavírus Em Bancos De Germoplasma De Inhame No Brasil. Summa Phytopathologica, 2020, 46:242-249.

Phillips, S., Briddon, R.W., Brunt, A.A., Hull, R. The Partial characterization of a badnavirus infecting the greater asiatic or water yam (*Dioscorea alata*). J. Phytopathol. 1999, 147:265-269.

Posada, D., Buckley, T. Model Selection And Model Averaging In Phylogenetics: Advantages Of Akaike Information Criterion And Bayesian Approaches Over LikelihoodRatio Tests. Systematic Biology, 2004, 53:793-808.

Rannala, B., Yang, Z.H. Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. J. Mol. Evol., 1996, 43:304-311.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst. Biol., 2012, 61:539-542.

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

Seal, S., Turaki, A., Muller, E., Kumar, P.L., Kenyon, L., Filloux, D., Galzi, S., Lopez-montes, A., Iskra-caruana, M. The prevalence of badnaviruses inWest African yams (*Dioscorea cayenensis-rotundata*) and evidence of endogenous pararetrovirus sequences in thier genomes. Virus Res. 2014, 186:144-154.

Silva, J.M., Jobim, L.J., Ramos-Sobrinho, R., Lima, J.S., Assunção, I.P., Cruz, M.M., Lima, G.S.A. Incidence and species diversity of badnaviruses infecting sugarcane from a

germplasm collection in Brazil. Trop. plant pathol., 2015, 40:212-217.

Sukal, A., Kidanemariam, D., Dale, J., James, A., Harding, R. 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 Res. 2017, 238:29-34.

Sukal, A.C., Kidanemariam, D.B., Dale, J.L., Harding, R.M., James, A.P. Characterization and genetic diversity of *Dioscorea bacilliform* viruses present in a Pacific yam germplasm collection. Plant Pathol. 2020, 69:576-584.

Thouvenel, J.C., Dumont, R. An Epidemiological Approach to The Study of Yam Mosaic Viruses In The Ivory Coast. Proc. Int. Soc. Tropical Root Crops, 1988, 45:643-649.

Thouvenel, J.C., Dumont, R. Yield decreases in yam infected with mosaic virus in Cote d'Ivoire. L'Agronomie Tropicale, 1990, 45:125-129.

Umber, M., Filloux, D., Muller, E., Laboureau, N., Galzi, S., Roumagnac, P., Pavis, C., Teycheney, P., Seal, S.E. The genome of African yam (*Dioscorea cayenensis-rotundata* complex) hosts endogenous sequences from four distinct badnavirus species. Mol. Plant Pathol. 2014, 15:790-801.

Umber, M., Gomez, R., Gelabale, S., Bonheur, L., Pavis, C., Teycheney, P. 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. Arch. Virol. 2016, 162:517-521.

Yang, I.C., Hafner, G.J., Dale, J.L., Harding, R.M. Genomic characterization of taro bacilliform virus. Arch. Virol. 2003, 148:937-949.

List of tables

Table 1. Percentages of pairwise nucleotide identity based on partial RT-RNase H sequences of the new isolates and badnaviral isolates retrieved from NCBI-GenBank.

	DBALV2	**DBESV**	**DBRTV2**	**DBSNV***	**DBSNV†**	**DBRTV3**	**DBRTV1**	**DBTRV***	**DBTRV†**	**eDBV9**	**TaBCHV**	**eDBV12***	**eDBV12†**	**DBBLV**	**DBALV***	**DBALV†**	**TaBV**
DBALV2	89.0-100.0																
DBESV	70.8-71.6	100.0															
DBRTV2	64.4-66.5	64.6-67.2	94.0-100.0														
DBSNV*	69.3-71.2	69.3-69.7	68.4-71.2	97.4													
DBSNV†	69.9-74.2	68.4-69.1	67.8-69.5	82.3-83.2	97.9-99.4												
DBRTV3	69.7-71.7	66.3-68.6	69.0-71.7	76.4-79.4	78.7-80.4	90.6-100.0											
DBRTV1	67.0-70.2	68.5-68.9	71.0-72.7	71.0-71.3	71.4-74.4	69.5-70.2	99.1-99.4										
DBTRV*	69.7-71.3	68.4	65.4-65.5	67.0-67.4	67.4-68.4	65.9-66.3	73.2-73.6	100.0									
DBTRV†	67.8-68.6	66.1-66.3	64.5-67.6	66.3-67.8	66.1-67.6	66.7-67.1	71.3-71.5	86.0-86.2	99.8								
eDBV9	64.2-70.4	68.6-69.2	66.9-68.2	68.6-69.4	68.8-70.8	67.8-69.0	72.9-73.5	82.8-83.6	83.7-84.7	95.5							
TaBCHV	65.0-68.9	65.5-67.0	65.7-68.9	64.2-66.7	65.5-69.1	64.6-66.7	67.0-69.1	63.9-65.2	63.9-65.6	65.7-67.6	88.4-99.3						
eDBV12*	65.3-69.2	66.9-67.8	67.6-69.2	65.1-65.7	64.3-65.5	65.1-65.9	71.9-72.3	68.8-69.4	66.0-66.7	68.2-68.8	69.8-70.2	95.7					
eDBV12†	65.4-69.6	67.9-68.3	65.0-67.4	65.6-66.4	65.8-67.2	64.9-66.6	70.3-71.1	66.0-66.6	62.2-63.0	68.3-69.1	68.0-68.9	88.6-89.2	99.4-100.0				
DBBLV	67.2-70.7	67.6-68.9	66.5-70.1	67.0-70.9	64.6-70.2	66.5-70.5	70.0-71.3	66.5-69.1	66.1-69.4	69.0-74.5	67.0-71.2	75.7-79.9	73.4-77.5	82.6-99.1			
DBALV*	65.7-71.3	69.3-71.0	65.4-68.2	68.2-71.0	67.2-71.8	67.4-70.0	71.0-73.4	69.9-71.2	66.7-70.0	70.8-72.9	68.4-70.0	73.6-76.1	72.7-75.2	73.2-77.7	85.2-100.0		
DBALV†	69.1-70.5	69.5-69.9	66.9-68.7	69.1-70.1	69.5-71.1	68.5-69.7	71.7-72.1	69.9	66.7-67.4	71.7-72.5	69.1-70.1	73.9-75.1	73.6-74.6	74.1-76.5	84.3-95.2	99.0	
TaBV	61.3-66.0	61.8-64.1	59.3-61.4	60.5-62.0	62.7-68.4	62.0-65.3	61.7-64.8	61.7-62.0	61.7-62.9	60.9-64.8	62.4-66.9	61.4-62.5	60.2-63.1	59.9-65.0	61.6-67.7	63.0-65.2	86.1-99.8

DBALV=dioscorea bacilliform AL virus; DBALV2=dioscorea bacilliform AL virus 2; DBESV=dioscorea bacilliform ES virus; DBRTV1=dioscorea bacilliform RT virus 1; DBRTV2=dioscorea bacilliform RT virus 2; DBRTV3=dioscorea bacilliform RT virus 3; DBSNV=dioscorea bacilliform SN virus; DBTRV=dioscorea bacilliform TR virus; DBBLV=dioscorea bacilliform BL vírus; TaBV=taro bacilliform virus; TaBCHV=taro bacilliform CH virus; eDBV9=dioscorea rotundata endogenous virus; eDBV12=dioscorea rotundata endogenous virus.

* Sequences retrieved from GenBank.

† Sequences reported in this study.

Table 2. Badnaviruses associated with *Dioscorea bulbifera* in Brazil.

Isolate	**Species**	**Acronym**	**Host**	**GenBank Accession #**
DBMT1	*Dioscorea bacilliform TR virus*	DBTRV	*Dioscorea bulbifera*	OM628722
DBMT2	*Dioscorea bacilliform TR virus*	DBTRV	*Dioscorea bulbifera*	OM628723
DBM9	*Dioscorea bacilliform SN virus*	DBSNV	*Dioscorea bulbifera*	OM628724
DBAM1	*Dioscorea bacilliform SN virus*	DBSNV	*Dioscorea bulbifera*	OM628725
DBMU	*Dioscorea bacilliform SN virus*	DBSNV	*Dioscorea bulbifera*	OM628726
DBJG	*Dioscorea bacilliform SN virus*	DBSNV	*Dioscorea bulbifera*	OM628727
DBM8	*Dioscorea bacilliform SN virus*	DBSNV	*Dioscorea bulbifera*	OM628728
DBT6	*Dioscorea bacilliform SN virus*	DBSNV	*Dioscorea bulbifera*	OM628729
DBC2	*Dioscorea bacilliform SN virus*	DBSNV	*Dioscorea bulbifera*	OM628730
DBM6	*Dioscorea bacilliform SN virus*	DBSNV	*Dioscorea bulbifera*	OM628731
DBJ2	*Dioscorea bacilliform AL virus*	DBALV	*Dioscorea bulbifera*	OM628732
DBJ3	*Dioscorea bacilliform AL virus*	DBALV	*Dioscorea bulbifera*	OM628733
DBT2	*Dioscorea rotundata* endogenous virus	eDBV12	*Dioscorea bulbifera*	OM628734
DBB1	*Dioscorea rotundata* endogenous virus	eDBV12	*Dioscorea bulbifera*	OM628735
DB21	*Dioscorea rotundata* endogenous virus	eDBV12	*Dioscorea bulbifera*	OM628736
DBT3	*Dioscorea rotundata* endogenous virus	eDBV12	*Dioscorea bulbifera*	OM628737
DBG	*Dioscorea rotundata* endogenous virus	eDBV12	*Dioscorea bulbifera*	OM628738
DBH	*Dioscorea bacilliform BL virus*	DBBLV	*Dioscorea bulbifera*	OM628739
DBB2	*Dioscorea bacilliform BL virus*	DBBLV	*Dioscorea bulbifera*	OM628740
DB31	*Dioscorea bacilliform BL virus*	DBBLV	*Dioscorea bulbifera*	OM628741
DB32	*Dioscorea bacilliform BL virus*	DBBLV	*Dioscorea bulbifera*	OM628742
DBM123	*Dioscorea bacilliform BL virus*	DBBLV	*Dioscorea bulbifera*	OM628743
DB11	*Dioscorea bacilliform BL virus*	DBBLV	*Dioscorea bulbifera*	OM628744
DBCU2	*Dioscorea bacilliform BL virus*	DBBLV	*Dioscorea bulbifera*	OM628745

DBCO4	*Dioscorea bacilliform BL virus*	DBBLV	*Dioscorea bulbifera*	OM628746
DBE	*Dioscorea bacilliform BL virus*	DBBLV	*Dioscorea bulbifera*	OM628747

List of figures

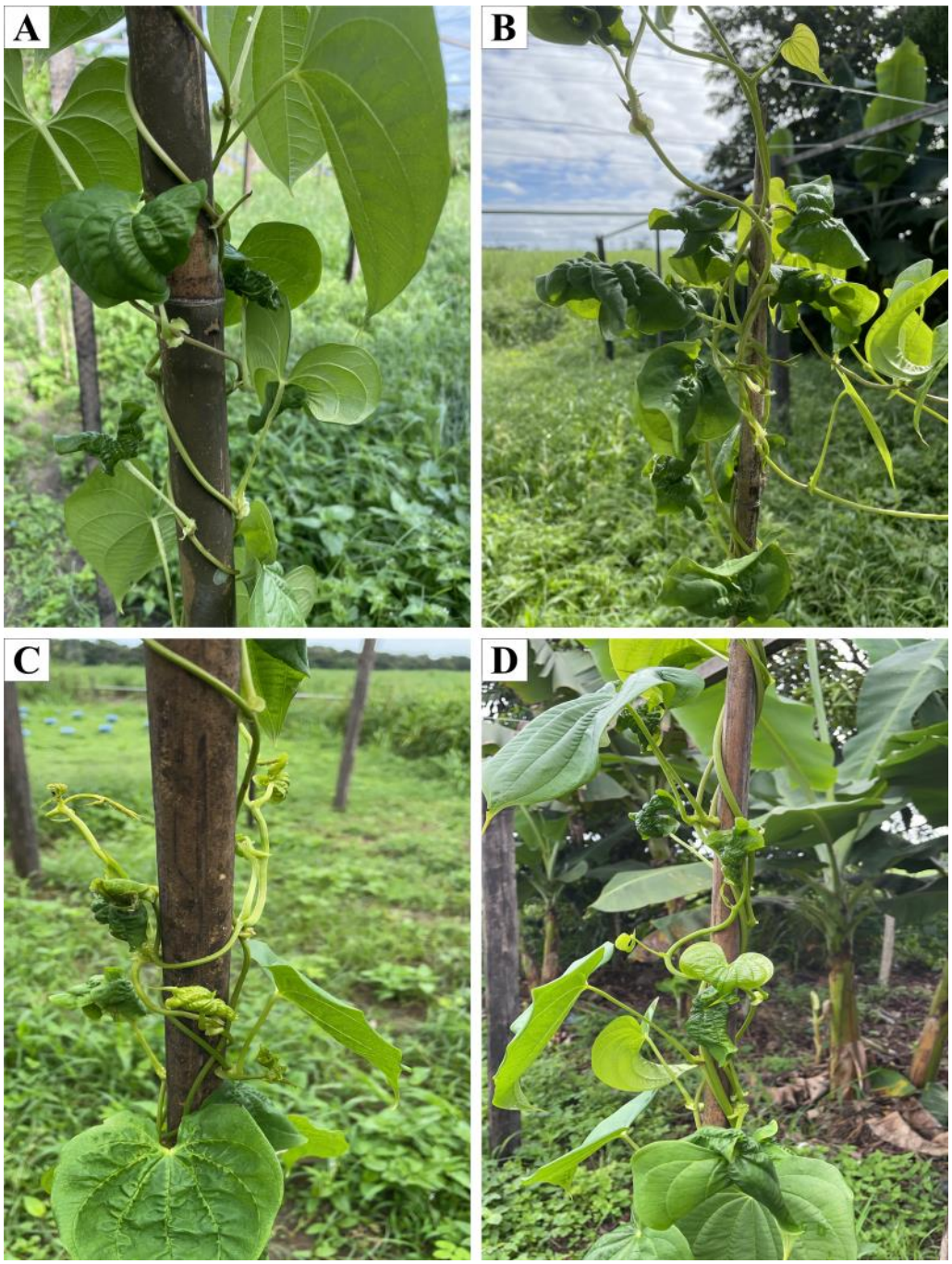


Figure 1. *Dioscorea bulbifera* plants growing at the experimental field of the Federal University of Alagoas, Alagoas state, Brazil, and exhibiting foliar chlorosis and leaf curling symptoms. Partial RT-RNase H sequences of *Dioscorea bacilliform SN virus* (A, isolate DBMU), *Dioscorea bacilliform TR virus* (B, isolate DBMT2), *Dioscorea bacilliform BL virus* (C, isolate DBM123), and *Dioscorea rotundata* endogenous virus (D, isolate DBB1) were characterized.

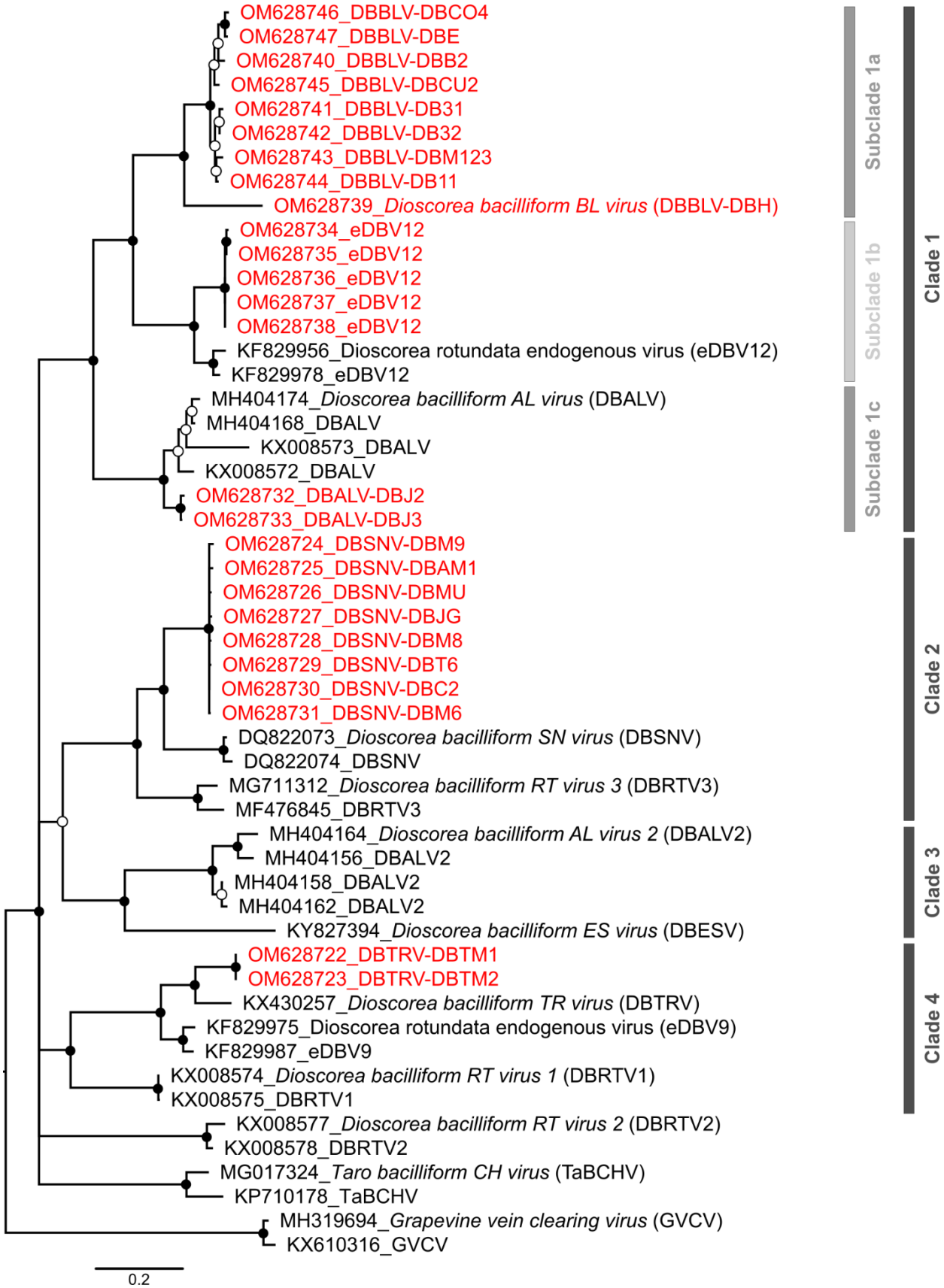


Figure 2. Bayesian phylogenetic tree based on partial RT-RNase H nucleotide sequences of badnaviruses. Posterior probability values between 0.95-1.0 (filled circles) and 0.50-0.94 (empty circles) are shown near to each branch node. The isolates reported here are indicated in red.

Supplementary Material

Supplementary Table S1. List of badnaviral sequences downloaded from NCBI-GenBank.

Species	Acronym	GenBank Accession #
Dioscorea bacilliform AL virus	DBALV	KX008571
Dioscorea bacilliform AL virus	DBALV	KX008572
Dioscorea bacilliform AL virus	DBALV	KX008573
Dioscorea bacilliform AL virus	DBALV	MG948562
Dioscorea bacilliform AL virus	DBALV	MH404165
Dioscorea bacilliform AL virus	DBALV	MH404166
Dioscorea bacilliform AL virus	DBALV	MH404167
Dioscorea bacilliform AL virus	DBALV	MH404168
Dioscorea bacilliform AL virus	DBALV	MH404169
Dioscorea bacilliform AL virus	DBALV	MH404170
Dioscorea bacilliform AL virus	DBALV	MH404171
Dioscorea bacilliform AL virus	DBALV	MH404172
Dioscorea bacilliform AL virus	DBALV	MH404173
Dioscorea bacilliform AL virus	DBALV	MH404174
Dioscorea bacilliform AL virus 2	DBALV2	MH404155
Dioscorea bacilliform AL virus 2	DBALV2	MH404156
Dioscorea bacilliform AL virus 2	DBALV2	MH404158
Dioscorea bacilliform AL virus 2	DBALV2	MH404159
Dioscorea bacilliform AL virus 2	DBALV2	MH404160
Dioscorea bacilliform AL virus 2	DBALV2	MH404161
Dioscorea bacilliform AL virus 2	DBALV2	MH404162
Dioscorea bacilliform AL virus 2	DBALV2	MH404163
Dioscorea bacilliform AL virus 2	DBALV2	MH404164
Dioscorea bacilliform AL virus 2	DBALV2	KY827395
Dioscorea bacilliform ES virus	DBESV	KY827394
Dioscorea bacilliform RT virus 1	DBRTV1	KX008574
Dioscorea bacilliform RT virus 1	DBRTV1	KX008575
Dioscorea bacilliform RT virus 1	DBRTV1	KX008576
Dioscorea bacilliform RT virus 2	DBRTV2	KX008577
Dioscorea bacilliform RT virus 2	DBRTV2	KX008578
Dioscorea bacilliform RT virus 2	DBRTV2	KY827393
Dioscorea bacilliform RT virus 2	DBRTV2	KX008579
Dioscorea bacilliform RT virus 3	DBRTV3	MG711311
Dioscorea bacilliform RT virus 3	DBRTV3	MG711312
Dioscorea bacilliform RT virus 3	DBRTV3	MF476845
Dioscorea bacilliform SN virus	DBSNV	DQ822073
Dioscorea bacilliform SN virus	DBSNV	DQ822074

Dioscorea bacilliform TR virus	DBTRV	KX430257
Taro bacilliform virus	TaBV	MG833013
Taro bacilliform virus	TaBV	MG017318
Taro bacilliform virus	TaBV	MG017322
Taro bacilliform virus	TaBV	MG017323
Taro bacilliform virus	TaBV	AF357836
Taro bacilliform CH virus	TaBCHV	MG017324
Taro bacilliform CH virus	TaBCHV	KP710177
Taro bacilliform CH virus	TaBCHV	KP710178
Dioscorea rotundata endogenous virus	eDBV9	KF829975
Dioscorea rotundata endogenous virus	eDBV9	KF829987
Dioscorea rotundata endogenous virus	eDBV12	KF829956
Dioscorea rotundata endogenous virus	eDBV12	KF829978

Supplementary Table S2. *Dioscorea bulbifera* plants collected throughout Brazil and tested, by PCR, using primers targeting the reverse transcriptase (RT) and ribonuclease H (RNase H) domains in *Badnavirus* genomes.

Sample code	**Municipality**	**State**	**PCR (BadnaFP/BadnaRP)***
DBG	Rio Largo	Alagoas	**Positive**
DBA	Mar Vermelho	Alagoas	Negative
DBJ1	Penedo	Alagoas	**Positive**
DBJ2	Penedo	Alagoas	**Positive**
DBJ3	Penedo	Alagoas	**Positive**
DBJ4	Penedo	Alagoas	Negative
DB11	Jaboatão dos Guararapes	Pernambuco	**Positive**
DB12	Jaboatão dos Guararapes	Pernambuco	**Positive**
DB21	Pombos	Pernambuco	**Positive**
DB22	Pombos	Pernambuco	Negative
DB31	Gravatá	Pernambuco	**Positive**
DB32	Gravatá	Pernambuco	**Positive**
DB51	Pombos	Pernambuco	Negative
DB52	Pombos	Pernambuco	Negative
DBT1	Palmeira dos Índios	Alagoas	Negative
DBT2	Palmeira dos Índios	Alagoas	**Positive**
DBT3	Palmeira dos Índios	Alagoas	**Positive**
DBT4	Palmeira dos Índios	Alagoas	**Positive**
DBT6	Palmeira dos Índios	Alagoas	**Positive**
DBT7	Palmeira dos Índios	Alagoas	Negative
DBB1	Belo Jardim	Pernambuco	**Positive**
DBB2	Belo Jardim	Pernambuco	**Positive**
DBJG	Joaquim Gomes	Alagoas	**Positive**
DBM1	Marechal Deodoro	Alagoas	Negative
DBM2	Marechal Deodoro	Alagoas	Negative
DBM3	Marechal Deodoro	Alagoas	**Positive**
DBM4	Marechal Deodoro	Alagoas	**Positive**
DBM5	Marechal Deodoro	Alagoas	**Positive**
DBM6	Marechal Deodoro	Alagoas	**Positive**
DBM8	Marechal Deodoro	Alagoas	**Positive**
DBM9	Marechal Deodoro	Alagoas	**Positive**
DBM10	Marechal Deodoro	Alagoas	Negative
DBE	Rio Largo	Alagoas	**Positive**
DBF	Flexeiras	Alagoas	Negative
DBH	Barra de Santo Antônio	Alagoas	**Positive**
DBM123	Murici	Alagoas	**Positive**
DBMU	Murici	Alagoas	**Positive**

DBMT1	Guiratinga	Mato Grosso	**Positive**
DBMT2	Guiratinga	Mato Grosso	**Positive**
DBC1	Rio Largo	Alagoas	**Positive**
DBC2	Rio Largo	Alagoas	**Positive**
DBAM1	Manaus	Amazonas	**Positive**
DBAM2	Manaus	Amazonas	**Positive**
DBAM3	Manaus	Amazonas	**Positive**
DBAM4	Manaus	Amazonas	Negative
DBCO1	Coruripe	Alagoas	Negative
DBCO2	Coruripe	Alagoas	**Positive**
DBCO3	Coruripe	Alagoas	**Positive**
DBCO4	Coruripe	Alagoas	**Positive**
DBVI1	Viçosa	Alagoas	Negative
DBVI2	Viçosa	Alagoas	Negative
DBVI3	Viçosa	Alagoas	**Positive**
DBVI4	Viçosa	Alagoas	Negative
DBMG1	São José da Lapa	Minas Gerais	Negative
DBMG2	São José da Lapa	Minas Gerais	Negative
DBMG3	São José da Lapa	Minas Gerais	Negative
DBCU1	Curitiba	Paraná	Negative
DBCU2	Curitiba	Paraná	**Positive**
DBCU3	Curitiba	Paraná	Negative
DBCU4	Curitiba	Paraná	**Positive**

* Primers described by Yang et al. (2003).

CAPÍTULO 3

COMPLETE GENOME SEQUENCES OF NEWLY DISCOVERED *Dioscorea bacilliform AL virus* AND *Dioscorea bacilliform SN virus* ISOLATES FROM *Dioscorea bulbifera* L. IN NORTHEASTERN BRAZIL

Giancarlo B.L. Santos, Mayra M.M. Ferro, Iraildes P. Assunção, Roberto Ramos-Sobrinho, Gaus S.A. Lima. Complete genome sequences of newly discovered *Dioscorea bacilliform AL virus* and *Dioscorea bacilliform SN virus* isolates from *Dioscorea bulbifera* L. in northeastern Brazil. European Journal of Plant Pathology, *submitted*.

Complete genome sequences of newly discovered *Dioscorea bacilliform AL virus* and *Dioscorea bacilliform SN virus* isolates from *Dioscorea bulbifera* L. in northeastern Brazil

Giancarlo B.L. Santos[1,2], Mayra M.M. Ferro[2], Iraildes P. Assunção[2], Roberto Ramos-Sobrinho[2*], Gaus S.A. Lima[2*]

[1]Campus Coruripe, Instituto Federal de Alagoas, 57230-000, Coruripe, Alagoas, Brazil
[2]Campus de Engenharias e Ciências Agrárias, Universidade Federal de Alagoas, 57100-000, Rio Largo, Alagoas, Brazil

***Corresponding authors:** R.R.S. (robertorsobrinho@gmail.com) and G.S.A.L. (gaus@ceca.ufal.br)

Abstract

To assess the DNA virome associated with air yam (*Dioscorea bulbifera* L.) plants displaying virus-like symptoms in northeastern Brazil in 2021, total DNA preparations from PCR positive samples (n=3) were used as template for viral genome amplification via rolling circle amplification and subjected to high-throughput sequencing. Three Illumina-based full-length badnaviral genomes were recovered from 980 to 30,668 reads and a coverage depth of 33 to 963x. The new genomes ranged from 7,208 to 7,420 bp in size and showed typical badnaviral genomic organization with three main open reading frames (ORFs 1-3). Based on pairwise nucleotide identity for the reverse transcriptase (RT) and ribonuclease H (RNase H) sequences, the newly discovered isolates were identified as dioscorea bacilliform AL virus (DBALV; n=1) and dioscorea bacilliform SN virus (DBSNV; n=2), at 85.1-86.6 and 82.2-83.9%, respectively. At least two recombination events were predicted in the DBSNV-DBH isolate, with DBALV- and DBSNV-like sequences identified as putative parents. To our knowledge, these are the first complete genome sequences of yam-infecting badnaviruses described in Brazil.

Keywords: *Badnavirus, Caulimoviridae*, Dioscoraceae, emergent disease.

1. Introduction

Viruses into the genus *Badnavirus* (family *Caulimoviridae*) have circular, double-stranded DNA (dsDNA) genomes of 7.0-9.2 kbp in size and encode at least three open reading frames (ORFs). The protein encoded by ORF1 is virion-associated (Cheng et al., 1996), while ORF2 encodes a nucleic acid-binding protein (Jacquot et al., 1996). The ORF3 represents almost 80% of the badnaviral genome and encodes a polyprotein including conserved domains such as viral capsid, movement protein, aspartate protease, reverse transcriptase (RT), and ribonuclease H (RNase H) (Medberry et al., 1990; Geering and Hull, 2012). Badnavirus genomes are encapsidated in a non-enveloped bacilliform particle, being mainly transmitted by mealybugs (Hemiptera: Pseudococcidae) in a semi-persistent manner. Badnaviruses are known as plant pararetroviruses because they replicate via an RNA intermediate, which also serves as template for protein translation (Geering and Hull, 2012; Bhat et al., 2016).

The genus *Dioscorea* (family Dioscoreaceae) comprises several socioeconomically important staple crops referred to as yams and are grown primarily by smallholder farmers in tropical and subtropical regions worldwide. Among the viral genera considered limiting factors to production of yams, *Badnavirus* has been reported associated with these host species in the main cultivation areas (Guimarães et al., 2015; Bhat et al., 2016; Sukal et al., 2020; Luo et al., 2022). Yam plants infected by a potyvirus, displaying internal brown spot disease symptoms, were first described to contain bacilliform particles during the 1970s in the Caribbean (Harrison and Roberts, 1973; Mantell and Haque, 1979). In the 1990s, molecular characterization of bacilliform viruses associated with *Dioscorea* spp. showed the presence of a new badnavirus tentatively named as dioscorea alata bacilliform virus (Phillips et al., 1999; Briddon et al., 1999). Currently, a complex of badnavirus species associated with yams is widespread throughout the world (Seal and Muller, 2007; Eni et al., 2008; Kenyon et al., 2008; Bousalem et al., 2009), and at least eight species are officially accepted by the International Committee on Taxonomy of Viruses (ICTV), *Dioscorea bacilliform AL virus* (previously known as dioscorea alata bacilliform virus), *Dioscorea bacilliform AL virus 2*, *Dioscorea bacilliform ES virus*, *Dioscorea bacilliform RT virus 1*, *Dioscorea bacilliform RT virus 2*, *Dioscorea bacilliform RT virus 3*, *Dioscorea bacilliform SN virus* (previously named as dioscorea sansibarensis bacilliform virus), and *Dioscorea bacilliform TR virus* (https://talk.ictvonline.org/ictv-reports).

Yam crops are of great relevance for food security in Brazil, with *D. alata* and *D. cayennensis* being the most commonly cultivated species. It has been shown that *Dioscorea bacilliform AL virus* (DBALV) seems to be the prevalent badnavirus associated with these hosts in the northeastern part of the country (Guimarães et al., 2015). Air yam (*Dioscorea bulbifera*

L.) is an edible crop native to both Asia and tropical Africa, and it is believed this species was introduced into the Americas probably during the slave trade from West Africa (Croxton et al., 2011; Ngwe et al., 2015). The badnaviruses DBALV and *Dioscorea bacilliform SN virus* (DBSNV) have been previously reported infecting *D. bulbifera* (Sukal et al., 2020; Nascimento et al., 2020). However, although the economic importance of yams, no badnaviral full-length genome has been reported from *Dioscorea* spp. in Brazil.

2. Materials and Methods

Plant samples and viral amplification

During 2021, air yam plants were collected from the state of Alagoas (AL), northeastern Brazil, and tested, by PCR, using a primer set amplifying the badnaviral RT-RNase H genomic region (Yang et al., 2003). To molecularly characterize the circular DNA virome associated with PCR-positive *D. bulbifera* samples from AL, total DNA preparations from three individual samples were used as template for rolling circle amplification (RCA) using the Illustra TempliPhi Amplification Kit (GE Healthcare, Buckinghamshire, United Kingdom) according to the manufacturer's instructions.

High-throughput sequencing

The RCA reactions were subjected to high-throughput sequencing (HTS) using the Illumina MiSeq platform at the Life Sciences Core Facility (LaCTAD) from State University of Campinas (UNICAMP) (Campinas, São Paulo, Brazil). The paired-end libraries (250 bp mean insert size) were constructed using the Illumina DNA Prep kit (Illumina Inc., California, USA) following the manufacturer's protocol.

De novo sequence assembly

The quality of Illumina reads was assessed using the FastQC v.0.11.5 program (www.bioinformatics.babraham.ac.uk/projects/fastqc/), and low-quality reads (Q score < 20) and adaptor sequences were trimmed using Trimmomatic v.0.38 (Bolger et al., 2014) with a sliding window of 4. The trimmed Illumina reads were *de novo* assembled in SPAdes v.3.15.4 (Bankevich et al., 2012). To additionally evaluate the quality of the circular contigs, trimmed reads were mapped against the badnavirus-like genome sequences using Bowtie2 (Langmead and Salzberg, 2012), and manually adjusted in IGV v.2.4.13 (Robinson et al., 2017). Finally, the Illumina-based contigs were analyzed using BLASTn (Altschul et al., 1990) to search for their more closely related virus sequences available in the NCBI-GenBank non-redundant

database. The predicted viral ORFs were identified using NCBI-ORF Finder (www.ncbi.nlm.nih.gov/orffinder).

Viral species demarcation and recombination analysis

The sequence dataset used here was comprised by the new isolates recovered from *D. bulbifera* and complete genomes of yam- and taro (*Colocasia esculenta*)-infecting badnaviruses downloaded from NCBI-GenBank (accessed on February 20, 2022; Supplementary Table S1). Further, the RT-RNase H (~1,230 bp) nucleotide sequences were retrieved from the full-length genomes. Both complete genome and RT-RNase H datasets were used for pairwise nucleotide sequence comparisons through the Sequence Demarcation Tool (SDT) v.1.2 (Muhire et al., 2014), and the ICTV-approved species demarcation criterion of ≥80% nucleotide identity for the RT-RNase H loci was adopted (Geering and Hull, 2012). To identify putative recombination breakpoints and parental sequences, the complete genome dataset was aligned using the MUSCLE algorithm (Edgar, 2004), manually edited in MEGA7 (Kumar et al., 2015), and subjected to recombination analysis in the Recombination Detection Program (RDP) v.4. (Martin et al., 2015). Recombination events detected by at least five different methods in the RDP4 package were considered reliable predictions.

Phylogenetic analysis

Bayesian phylogenetic trees based on complete genome and RT-RNase H nucleotide sequences were inferred through the CIPRES web portal (Miller et al., 2010) using MrBayes v.3.2.3 (Ronquist et al., 2012), assuming a general time reversible (GTR) nucleotide substitution model with a gamma (G) model of rate heterogeneity and invariable (I) sites, determined using MrModeltest v.2.3 (Posada and Buckley, 2004) according to the Akaike Information Criterion (AIC). The analyses consisted of two replicates with four chains each for 10 million generations and sampling every 1,000 generations. The first 2,500 trees per run were discarded as a burn-in. The posterior probability values (Rannala and Yang, 1996) were determined from the majority-rule consensus tree reconstructed with the 15,000 remaining trees. The phylogenetic trees were visualized and edited in FigTree v.1.4 (ztree.bio.ed.ac.uk/software/figtree) and Inkscape v.1.1 (https://inkscape.org/pt/).

3. Results and Discussion

Three newly discovered badnaviral isolates were characterized, at the genome level, from air yam plants. Based on the ≥80% nucleotide identity threshold for *Badnavirus* species demarcation, the SDT analysis showed that the RT-RNase H sequences from these isolates

belong to DBALV (n=1) and DBSNV (n=2), at 85.1-86.6 and 82.2-83.9%, respectively, with complete genomes sharing 83.4-85.6 and 78.6-80.5% identity (Supplementary Figures S1 and S2). The two new DBSNV genomes showed 89.2% nucleotide identity one with another (Supplementary Figure S2). The Illumina-based genomes were 7,420 (DBALV-DBMU), 7,208 (DBSNV-DBJG), and 7,228 (DBSNV-DBH) bp in length (Fig. 1), being derived from 30,668, 28,824, and 980 reads and a coverage depth of 963, 932, and 33x, respectively. The ORFs 1-3 were predicted from each viral genome, and additional badnaviral hallmark features such as $tRNA^{met}$ primer binding site sequence, TATA box, and polyadenylation signal were found (Fig. 1). The new DBALV and DBSNV genome sequences were submitted to the NCBI-GenBank database under accession nos. ON402786-ON402788.

The RDP4 package identified two independent recombination events involving the DBSNV-DBH isolate (Supplementary Table S2). The recombination breakpoints of event 1 were located in ORF1, with DBALV-DBMU and DBSNV-DBJG predicted as potential minor and major parents, respectively (Fig. 1; Supplementary Table S2). The event 2 was identified in ORF3 and DBSNV-like isolates were predicted as the parental sequences, suggesting it might be an intraspecies recombination event (Fig. 1; Supplementary Table S2). Both recombination events were identified by at least five different methods implemented in RDP4, and the highest significant *P* value was 3.9×10^{-3} (Supplementary Table S2).

The Bayesian phylogenetic tree based on complete genome sequences showed the close genetic relationship among the DBALV and DBSNV isolates reported in the present study and sequences retrieved from NCBI-GenBank, with these badnaviral species forming two different phylogenetic groups (Fig. 2a). The phylogenetic tree based on RT-RNase H nucleotide sequences displayed some topological incongruences when compared to full-length genomes. The DBSNV and *Dioscorea bacilliform RT virus* 3 (DBRTV3) isolates grouped into two sister clades in the complete genome tree, but they seem to comprise a polyphyletic group in the RT-RNase H tree (Fig. 2a and 2b). However, it is important to note that the value supporting this putative polyphyletic clustering (posterior probability = 0.76) is much lower than the support for DBSNV and DBRTV3 forming sister clades (posterior probability = 1.0) in the complete genome tree (Fig. 2). These results suggest the RT-RNase H dataset has insufficient phylogenetic signal to resolve the genetic relationship among *Dioscorea*-infecting badnaviruses known thus far.

High badnaviral species diversity has been reported associated with yams worldwide, and the eight ICTV-approved species have been molecularly characterized at the genome level (Briddon et al., 1999; Seal and Muller, 2007; Bomer et al., 2016; Umber et al., 2016; Sukal et al., 2017; Bomer et al., 2018; Sukal et al., 2020). Three Illumina-determined DBALV complete

genomes have been recently described from air yam germplasm materials kept at the Centre for Pacific Crops and Trees (CePACT) in Fiji (Sukal et al., 2020). Also, partial DBSNV RT-RNase H sequences have been recovered from an air yam accession of the gene bank collection of the University of São Paulo, Brazil (Nascimento et al., 2020). Here, *D. bulbifera*-infecting DBALV and DBSNV complete genome sequences were characterized from commercially grown plants in AL, northeastern Brazil. Together, these findings reinforce the importance of establishing reliable diagnostic tools for detection and identification of yam-infecting badnaviruses that should be used in routine viral screenings to avoid the spread of infected materials.

Recombination has been reported as an important evolutionary mechanism acting on genetic diversification of badnaviruses infecting economically important crops such as cacao (*Theobroma cacao*) and banana (*Musa* spp.) (Sharma et al., 2015; Ramos-Sobrinho et al., 2020). The DBRTV3 isolate originally described from *D. rotundata* has been identified as a recombinant sequence having DBALV and DBSNV isolates as putative parents (Bomer et al., 2018). Here, the DBSNV-DBH isolate was implicated in at least two well-supported recombination events, with DBALV and DBSNV also being predicted as the isolates more closely related to the parental sequences. Together, these results suggest that recombination might play a role on the molecular diversity observed in yam-infecting badnaviruses.

References

Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410.

Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.; Nikolenko, S.; Pham, S.; Prjibelski, A.; Pyshkin, A.; Sirotkin, A.; Vyahhi, N.; Tesler, G.; Alekseyev, M.A.; Pevzner, P.A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* 19(5):455-77.

Bhat, A.I.; Hohn, T.; Selvarajan, R. (2016). Badnaviruses: The Current Global Scenario. *Viruses* 177, 1–29.

Bolger, A.M.; Lohse, M.; Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* 30(15): 2114–2120.

Bömer, M.; Rathnayake, A.I.; Visendi, P.; Silva, G.; 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.; Turaki, A.; Silva, G.; Kumar, P.; Seal, S. (2016). A sequence-independent strategy

for amplification and characterization of episomal badnavirus sequences reveals three previously uncharacterized yam badnaviruses. *Viruses* 8:188.

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.

Cheng, C.P.; Lockhart, B.E.; Olszewski, N.E. (1996). The ORF I and II Proteins of *Commelina yellow mottle virus* are virion-associated. *Virology* 223:263-271.

Croxton, M.D.; Andreu, M.A.; Williams, D.A.; Overholt, W.A.; Smith, J.A. (2011). Geographic Origins and Genetic Diversity of Air-Potato (*Dioscorea bulbifera*) in Florida. *Invasive Plant Science and Management* 4:22-30.

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, 1792–1797.

Eni, A.O.; Hughes, J.D.A.; Asiedu, R.; 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.

Geering, A.D.W.; Hull, R. (2012). Family Caulimoviridae. In Virus Taxonomy. 9th Report of the International Committee on Taxonomy of Viruses; King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., Eds.; Elsevier Acad. Press: London, UK, 429–443.

Guimarães, K.M.; Silva, S.J.C.; Melo, A.M.; Ramos-Sobrinho, R; Lima, J.S.; Zerbini, F.M.; Assunção, I.P., Lima, G.S.A. (2015) Genetic variability of badnaviruses infecting yam (*Dioscorea* spp.) in northeastern Brazil. *Tropical Plant Pathology* 40:111-118.

Harrison, B.; Roberts, I. (1973). Association of virus-like particles with internal brown spot of yam (*Dioscorea alata*). *Tropical Agriculture* 50:335-340.

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

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.

Kumar, S.; Stecher, G.; Tamura, K. (2015). MEGA7: Molecular Evolutionary Genetics Analysis version 7. *Molecular Biology Evolution* 33, 1870–1874, doi:10.1093/molbev/msw054.

Langmead B.; Salzberg S. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9:357-359.

Luo, G.F.; Podolyan, A.; Kidanemariam, D.B.; Pilotti, C.; Houliston, G.; Sukal, A.C. (2022). A Review of Viruses Infecting Yam (*Dioscorea* spp.). *Viruses* 14:662.

Mantell, S.H.; Haque, S.Q. Internal brown spot disease of yams. Yam Virus Proj. Bull. (Trinidad Tobago), ill, 1979, 1–13.

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

Medberry, S.L.; Lockhart, B.E.L.; Olszewski, N.E. (1990). Properties of *Commelina yellow mottle virus's* complete *DNA* sequence, genomic discontinuities and transcript suggest that it is a pararetrovirus. *Nucleic Acids Research* 18:5505-5513.

Miller, M.A.; Holder, M.T.; Vos, R.; Midford, P.E.; Liebowitz, T.; Chan, L.; Hoover, P.; Warnow, T. The CIPRES Portals. 2010. CIPRES. Available online: http://www.phylo.org/sub-sections/portal (accessed on March 12, 2022).

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

Nascimento, J.P.; Oliveira, M.L.; Peixinho, G.S.; Ferro, M.M.M.; Silva, S.J.C.; Versey, E.A.; Lima, G.S.A.; Assunção, I.P. (2020). Incidência E Caracterização Molecular De Badnavírus Em Bancos De Germoplasma De Inhame No Brasil. *Summa Phytopathologica* 46:242-249.

Ngwe, M.F.S.N.; Omokolo, D.N.; Joly, S. (2015) Evolution and Phylogenetic Diversity of Yam Species (*Dioscorea* spp.): Implication for Conservation and Agricultural Practices. *PLoS ONE* 10(12): e0145364. https://doi.org/10.1371/journal.pone.0145364

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

Posada, D.; Buckley, T. (2004). Model Selection And Model Averaging In Phylogenetics: Advantages Of Akaike Information Criterion And Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology* 53:793-808.

Ramos-Sobrinho, R.; Chingandu, N.; Gutierrez, O.A.; Marelli, J.-P.; and Brown, J.K. (2020). A complex of badnavirus species infecting cacao reveals mixed infections, extensive genomic variability, and interspecific recombination. *Viruses* 12, 443; doi:10.3390/v12040443.

Rannala, B.; Yang, Z.H. (1996). Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *Journal of Molecular Evolution* 43:304-311.

Robinson, J.T.; Thorvaldsdóttir, H.; Wenger, A.M.; Zehir, A.; Mesirov, J.P. (2017). Variant

Review with the Integrative Genomics Viewer (IGV). *Cancer Research* 77:31-34

Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Hohna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematics Biology* 61:539-542.

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

Sharma, S.K.; Vignesh-Kumar, P.; Geetanjali, A.S.; Pun, K.B.; Baranwal, V.K. (2015). Subpopulation level variation of banana streak viruses in India and common evolution of banana and sugarcane badnaviruses. *Virus Genes* 50(3):450-65.

Sukal, A.; Kidanemariam, D.; Dale, J.; James, A.; 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.; James, A.P. (2020). Characterization and genetic diversity of *Dioscorea bacilliform* viruses present in a Pacific yam germplasm collection. *Plant Pathology* 69:576-584.

Umber, M.; Gomez, R.; Gelabale, S.; Bonheur, L.; Pavis, C.; Teycheney, P. (2016). 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. *Archives of Virology* 162:517-521.

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

List of figures

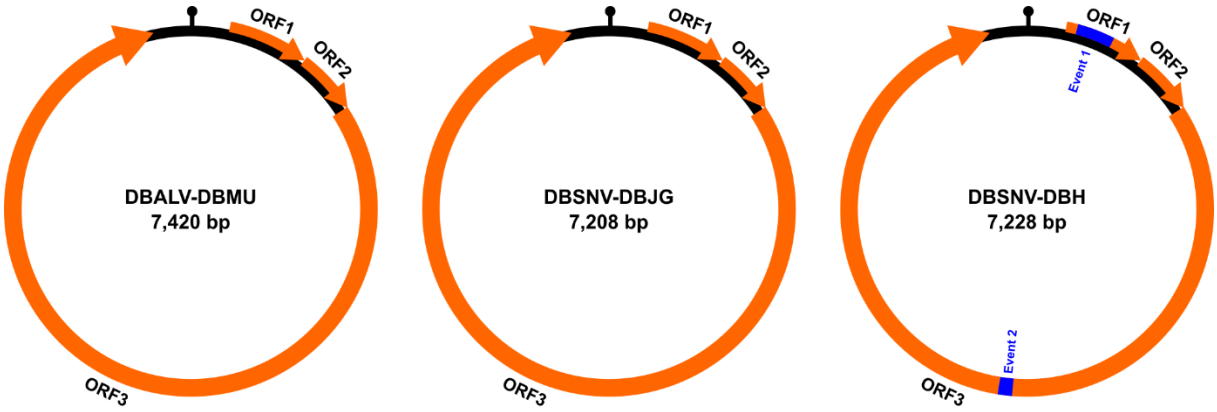


Fig. 1. Genomic organization of the newly discovered badnaviruses dioscorea bacilliform AL virus (DBALV-DBMU) and dioscorea bacilliform SN virus (DBSNV-DBJG and DBSNV-DBH) infecting *Dioscorea bulbifera* in Brazil. The recombinant regions in DBSNV-DBH, detected by the RDP4 package, are indicated by blue boxes [event 1 (minor parent – DBALV-DBMU; major parent – DBSNV-DBJG) and event 2 (minor parent – DBSNV-DQ822073; major parent – DBSNV-DBJG)].

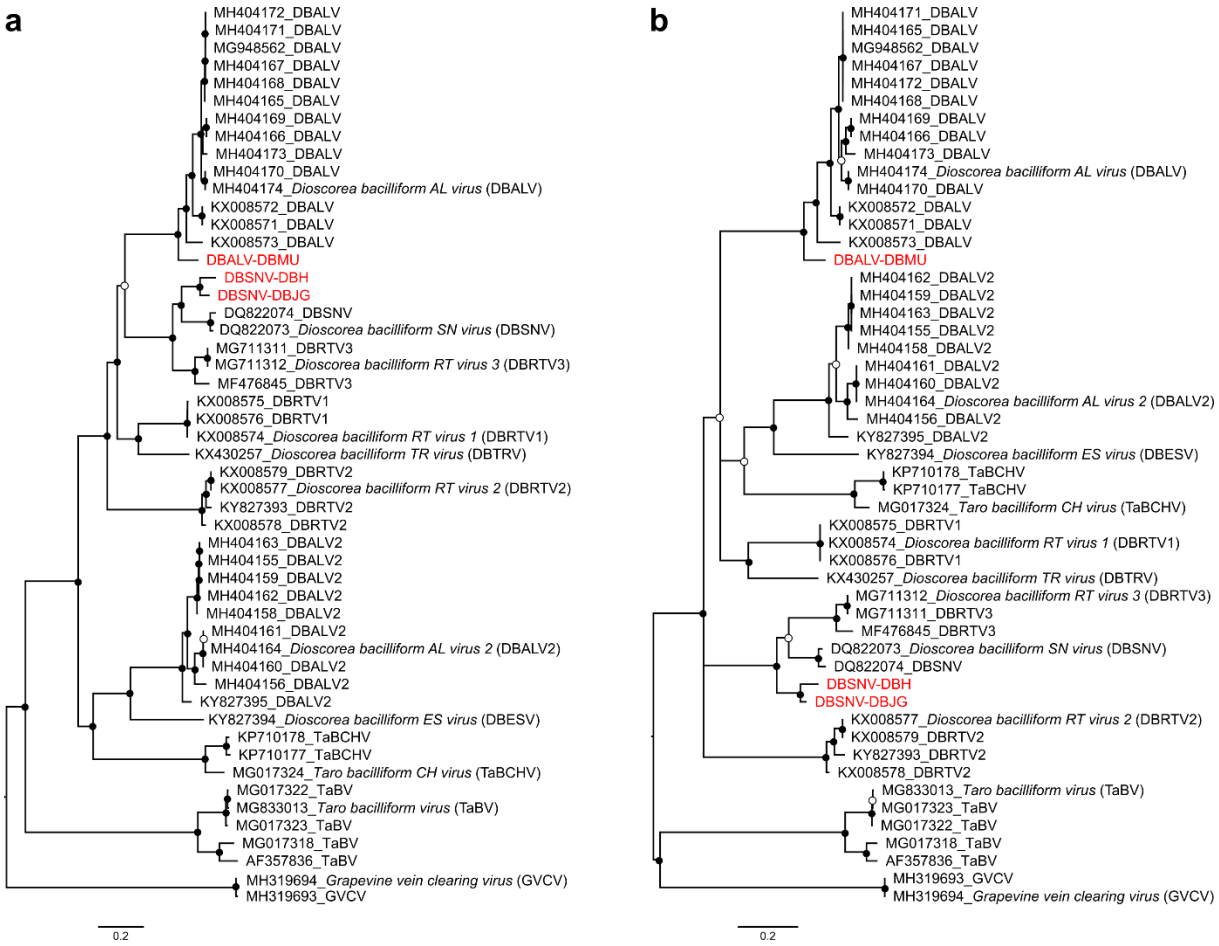


Fig. 2. Bayesian phylogenetic trees based on complete genomes (a) and RT-RNase H (b) nucleotide sequences of yam-infecting badnaviruses. Posterior probability values between 0.99-1.0 (filled circles) and 0.50-0.98 (empty circles) are shown near to each branch node. The isolates reported here are indicated in red. Dioscorea bacilliform AL virus (DBALV), dioscorea bacilliform AL virus 2 (DBALV2), dioscorea bacilliform RT virus 1 (DBRTV1), dioscorea bacilliform RT virus 2 (DBRTV2), dioscorea bacilliform RT virus 3 (DBRTV3), dioscorea bacilliform TR virus 1 (DBTRV), dioscorea bacilliform SN virus (DBSNV), dioscorea bacilliform ES virus (DBESV), taro bacilliform virus (TaBV), taro bacilliform CH virus (TaBCHV), and grapevine vein clearing virus (GVCV).

Supplementary Material

Supplementary Table S1. Complete genome sequences of yam-infecting badnaviruses downloaded from NCBI-GenBank (accessed on February 20, 2022).

Species	Acronym	GenBank Accession Nos.
Dioscorea bacilliform AL virus	DBALV	MG948562, MH404165, MH404166, MH404167, MH404168, MH404169, MH404170, MH404171, MH404172, MH404173, MH404174, KX008571, KX008572, KX008573
Dioscorea bacilliform AL virus 2	DBALV2	KY827395, MH404155, MH404156, MH404158, MH404159, MH404160, MH404161, MH404162, MH404163, MH404164
Dioscorea bacilliform RT virus 1	DBRTV1	KX008574, KX008575, KX008576
Dioscorea bacilliform RT virus 2	DBRTV2	KY827393, KX008577, KX008578, KX008579
Dioscorea bacilliform RT virus 3	DBRTV3	MG711311, MG711312, MF476845
Dioscorea bacilliform TR virus	DBTRV	KX430257
Dioscorea bacilliform SN virus	DBSNV	DQ822073, DQ822074
Dioscorea bacilliform ES virus	DBESV	KY827394
Taro bacilliform virus	TaBV	AF357836, MG017322, MG017323, MG833013, MG017318
Taro bacilliform CH virus	TaBCHV	KP710178, KP710177, MG017324
Grapevine vein clearing virus	GVCV	MH319693, MH319694

Supplementary Table S2. Predicted recombination events detected, based on complete genome sequences, within the new badnavirus isolates infecting *Dioscorea bulbifera*.

	Breakpoints[1]			**Parents**			***P* value**[3]	
Event	**Begin**	**End**	**Recombinant**	**Minor**	**Major**	**Methods**[2]	**Highest**	**Lowest**
1	183	629	^DBSNV-DBH	DBALV-DBMU	DBSNV-DBJG	RBM**C**3	3.91E-03	9.63E-10
2	3637	3784	^DBSNV-DBH	DBSNV-DQ822073	DBSNV-DBJG	RGBM**C**3	4.19E-03	1.47E-07

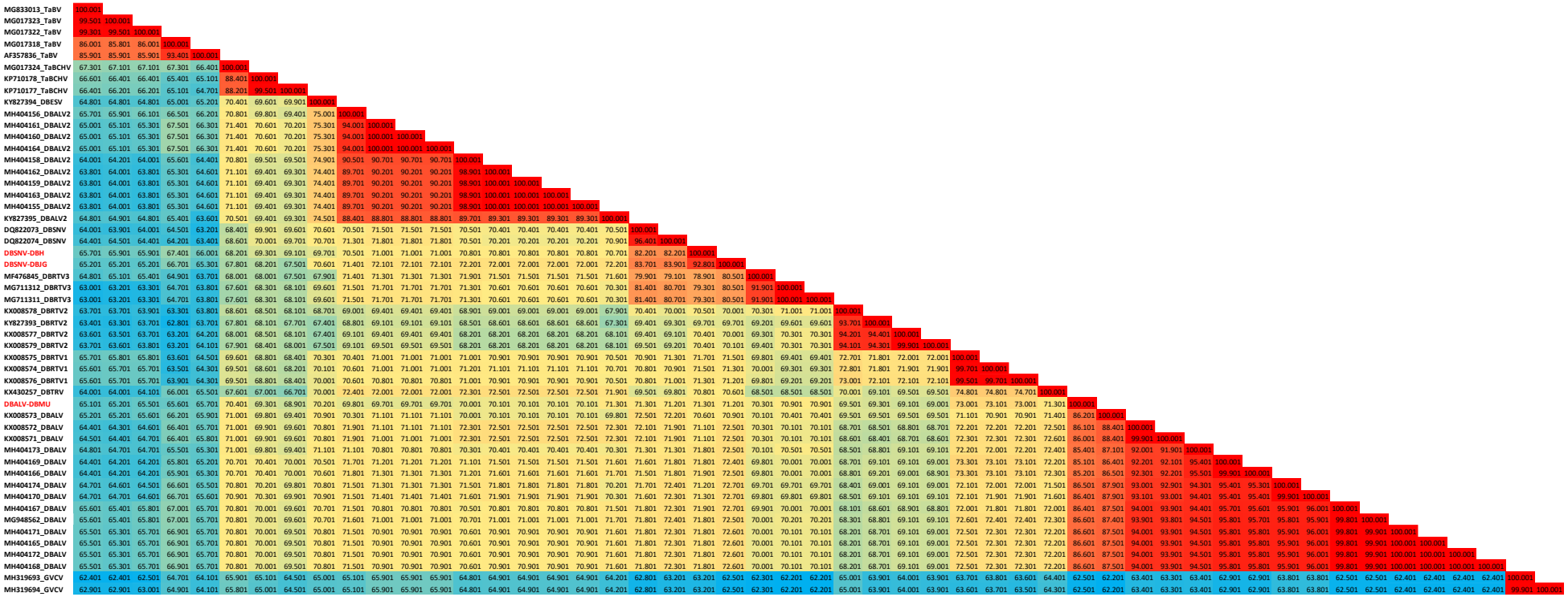
[1]Numbering starts at the 5' end of the minus-strand primer-binding site and increases clockwise.

[2]R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimera; S, SisScan; 3, 3SEQ.

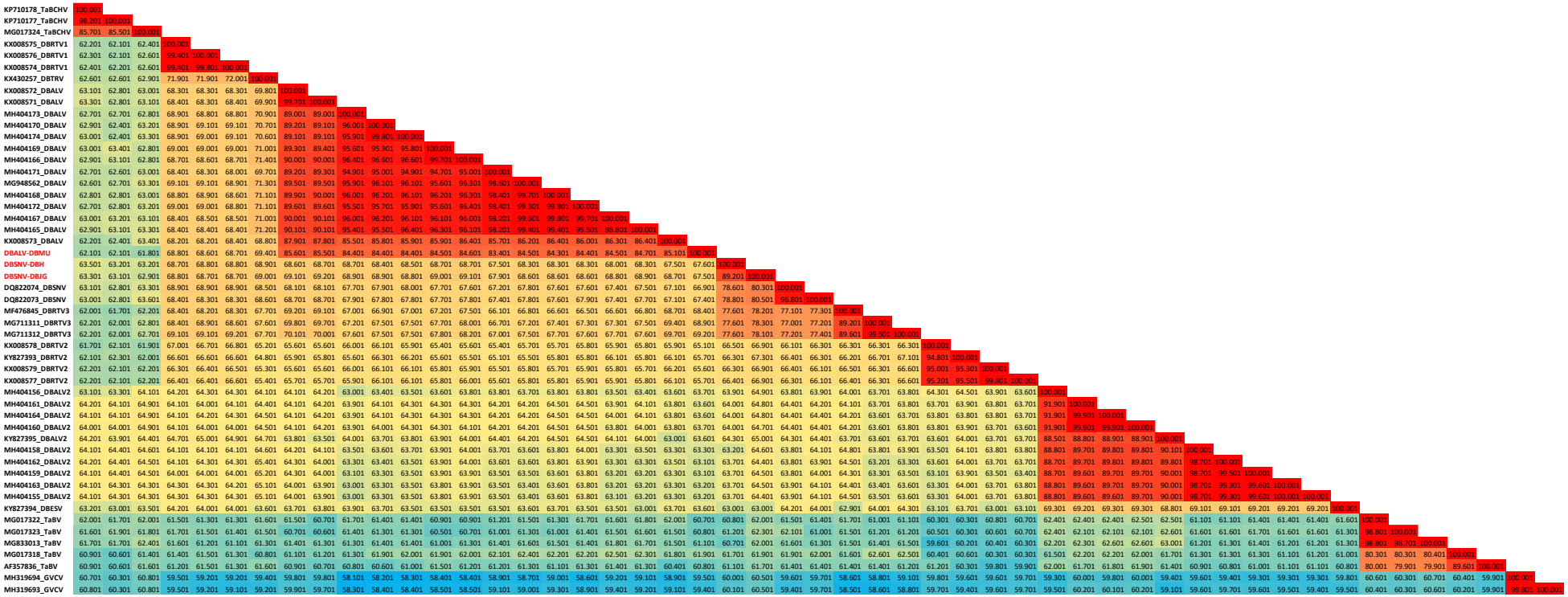
[3]The *P* values are for the methods indicated in bold (highest) and red (lowest).

^ The recombinant sequence may have been misidentified (one of the identified parents might be the recombinant).

DBSNV=dioscorea bacilliform SN virus; DBALV= dioscorea bacilliform AL virus.



Supplementary Figure S1. Pairwise sequence comparisons of RT-RNase H nucleotide sequences of yam-infecting badnaviruses. Dioscorea bacilliform AL virus (DBALV), dioscorea bacilliform AL virus 2 (DBALV2), dioscorea bacilliform RT virus 1 (DBRTV1), dioscorea bacilliform RT virus 2 (DBRTV2), dioscorea bacilliform RT virus 3 (DBRTV3), dioscorea bacilliform TR virus 1 (DBTRV), dioscorea bacilliform SN virus (DBSNV), dioscorea bacilliform ES virus (DBESV), taro bacilliform virus (TaBV), taro bacilliform CH virus (TaBCHV), and grapevine vein clearing virus (GVCV). The new isolates are indicated in red.



Supplementary Figure S2. Pairwise sequence comparisons of complete genomes of yam-infecting badnaviruses. Dioscorea bacilliform AL virus (DBALV), dioscorea bacilliform AL virus 2 (DBALV2), dioscorea bacilliform RT virus 1 (DBRTV1), dioscorea bacilliform RT virus 2 (DBRTV2), dioscorea bacilliform RT virus 3 (DBRTV3), dioscorea bacilliform TR virus 1 (DBTRV), dioscorea bacilliform SN virus (DBSNV), dioscorea bacilliform ES virus (DBESV), taro bacilliform virus (TaBV), taro bacilliform CH virus (TaBCHV), and grapevine vein clearing virus (GVCV). The new isolates are indicated in red.