



**UNIVERSIDADE FEDERAL RURAL DE
PERNAMBUCO**

Pró-Reitoria de pesquisa e pós-graduação



**PROGRAMA DE PÓS-GRADUAÇÃO
EM FITOPATOLOGIA**

Tese de Doutorado

**Análise genômica de potyvírus associados a espécies de
dioscoreáceas no Brasil**

Alejandro Risco Mendoza

**Recife – PE
2022**

ALEJANDRO RISCO MENDOZA

**ANÁLISE GENÔMICA DE POTYVÍRUS ASSOCIADOS A ESPÉCIES DE
DIOSCOREÁCEAS NO BRASIL**

Tese apresentada ao Programa de Pós-Graduação em Fitopatologia da Universidade Federal Rural de Pernambuco, como parte dos requisitos para obtenção do título de Doutor em Fitopatologia.

COMITÊ DE ORIENTAÇÃO:

Profa. Dra. Rosana Blawid (UFRPE) – Orientador

**RECIFE-PE
FEVEREIRO – 2022**

**ANÁLISE GENÔMICA DE POTYVÍRUS ASSOCIADOS A ESPÉCIES DE
DIOSCOREÁCEAS NO BRASIL**

ALEJANDRO RISCO MENDOZA

Tese ----- e -----pela Banca Examinadora em: 21 de fevereiro de 2022

ORIENTADOR:

Prof^a. Dr^a. Rosana Blawid
(UFRPE)

EXAMINADORES:

Prof^a. Dr^a. Elineide Barbosa de Souza (UFRPE)

Prof. Dr. Leonardo Silva Boiteux (UFRPE)

Prof. Dr. Raúl Humberto Blas Sevillano (UMALM, Perú)

Prof. Dr. Tatsuya Nagata (UnB)

**RECIFE-PE
FEVEREIRO – 2022**

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

R595a Risco Mendoza, Alejandro
 Análise genômica de potyvírus associados a espécies de dioscoreáceas no Brasil / Alejandro Risco
Mendoza. - 2022.
 172 f. : il.

 Orientador: Profa. Dra. Rosana Blawid.
 Inclui referências.

 Tese (Doutorado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em
Fitopatologia, Recife, 2022.

 1. Yam mosaic virus. 2. Yam mild mosaic virus. 3. Dioscorea spp.. 4. HTS. 5. Estudos
filogeográficos e geoespaciais de YMV. I. , Profa. Dra. Rosana, orient. II. Título

CDD 632

Dedico esta tese:

À minha família, minha mãe Marciana Mendoza de La Cruz; irmãos Edith, Marlene, Martha, Margarita, José e Iván; e a meus sobrinhos Naomi, Mía, José Manuel, Valentina, Maya, Jana, Santiago e Daniel.

AGRADECIMENTOS

Agradeço a Deus por sempre guiar meu caminho e me dar sabedoria quando preciso.

Agradeço a minha querida mãe que fez o papel de pai e mãe ao longo da minha vida, a meu pai Diógenes, sei que está sempre me protegendo e guiando meu caminho e aos meus irmãos, porque eles estão sempre dispostos a me dar seu apoio em todos os momentos e onde quer que eu vá.

À minha orientadora Profa. Dra. Rosana Blawid, que é uma mulher que luta incansavelmente por equidade e justiça; a ela que, desde o primeiro dia de permanência na UFRPE, me deu sua confiança, seu apoio, e que me ensinou sobre o mundo da virologia vegetal com muita paciência.

As instituições financiadoras desta pesquisa: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Edição Universal/2018, Universidade Federal Rural de Pernambuco (PRPPG Edição 015/2018), professor Doutor do Dep. Biologia Celular-UnB Tatsuya Nagata e Departamento de Vírus de Plantas do Instituto Leibniz DSMZ-Alemanha.

A todos os professores do Programa de Pós-Graduação em Fitopatologia da Universidade Federal Rural de Pernambuco, que contribuíram para minha formação no doutorado.

À Universidad Nacional Agraria La Molina, Lima, Perú, em especial ao departamento de Fitopatologia. Graças ao apoio das autoridades e professores, pude iniciar e terminar este doutorado.

Ao Programa de Bolsas Brasil-Cooperação entre a Organização dos Estados Americanos (OEA) e o Grupo Coimbra de Universidades Brasileiras (CGUB).

A todos os pequenos produtores de inhame do Nordeste do Brasil, que foram gentis e prestativos em fornecer amostras de inhame para a realização deste trabalho de pesquisa.

Aos profissionais, pelo fornecimento de amostras (túberas de inhame) e apoio nas coletas de amostras em áreas não cultivadas: Mestre Gabriel Pavan Sabino (UNESP), Dr. Diogo Amorim de Araújo, Pernambuco e o Dr. Jones Clebson R. Mendez (UFRPE).

Aos pesquisadores do Instituto Agrônomo de Pernambuco (IPA) Dr. José Nildo Tabosa, e ao MSc. Almir Dias Alves da Silva por nos dar acesso às áreas de produção de Inhame.

Agradecimentos especiais aos membros da LAFSOL por seu apoio contínuo. A Dra. Iwanne Coelho e os doutorandos Marcelo Garcia, Tiago Torres e Thais Almeida.

À equipe do laboratório de Fitovirologia da UFRPE, em especial a minha amiga Tatiane, Géssyka, Carlos Machado e Ailton, que me acompanharam neste processo de formação e me deram o seu apoio incondicional.

Aos moradores da República da Várzea, que tive a oportunidade de conhecê-los e conviver durante os quatro anos de minha permanência no Brasil: Ana, Renata, Eduarda, Octávio, Valdemir, Steven, Pedro, Laura, Arturene, Islaine, Ruan, David, Eloise, Indira, Sônia, Elias e Thiago.

Aos meus amigos: Juan José, Cledy Sierra, Ana Cristina Ramos e Ramires Eloise que me acompanharam nesta jornada.

SUMÁRIO

	página
RESUMO GERAL	ix
GENERAL ABSTRACT	x
CAPÍTULO I	12
INTRODUÇÃO GERAL	13
1. A cultura do inhame: aspectos nutricionais e econômicos	13
2. Viroses na cultura do inhame	16
3. Família <i>Potyviridae</i>	20
3.1. Organização genômica, expressão gênica e função das proteínas virais do gênero <i>Potyvirus</i>	31
3.2. Transmissão.....	39
3.3. Transmissão por afídeos.....	41
3.4. Sintomas e sinais	43
4. Controle	48
5. Sequenciamento de alto desempenho (<i>High-Throughput Sequencing</i> , HTS)	51
5.1. Plataformas HTS	52
5.2. Aplicações de HTS na detecção de vírus de plantas	55
5.3. Análise de dados HTS	56
6. Filogenia molecular	60
7. Relógio molecular	63
8. Referências Bibliográficas.....	67
CAPÍTULO II	101
Characterization of yam mosaic viruses from Brazil reveals a new phylogenetic group and possible incursion from the African continent.....	102
Abstract.....	102
1. Introduction	103
2. Material and Methods	104
2.1. Field surveys	104
2.2. High-Throughput Sequencing	104
2.3. Virome analysis.....	107
2.4. Bayesian inferences.....	108
2.5. Geospatial analysis	111
2.6. Evolutionary and temporal analysis	111
3. Results	113
3.1. Symptoms and serological detection.....	113
3.2. Virome analysis and characterization of YMV genomes.....	114
3.3. Pairwise sequence analyses of the Brazilian genomes.....	118
3.4. Bayesian analyses.....	120

3.5. Geospatial analysis.....	125
3.6. Evolutionary and temporal analysis	126
4. Discussion.....	129
References	133
CAPÍTULO III	139
Near-complete genomic sequences of yam mild mosaic viruses of <i>D. alata</i> from Brazil	140
1. Introduction	140
2. Material and methods	142
2.1. Sample collection and ELISA survey	142
2.2. RNA extraction and High-Throughput Sequencing.....	143
2.3. Viral genome assembly and sequence analysis.....	143
2.4. Bayesian inferences.....	144
2.5. Genomic sequence recombination analysis.....	149
3. Results and Discussion	149
3.1. Symptoms and ELISA results	149
3.2. Characterization of YMMV genomic sequences	151
3.3. Pairwise sequence analyses of the Brazilian genomes.....	155
3.4. Bayesian analyses.....	157
3.5. Genomic recombination events in YMMV	160
4. Conclusions	162
References	164
CONCLUSÕES GERAIS	171

RESUMO GERAL

O inhame (*Dioscorea* L., Família Dioscoreaceae) é uma cultura que se desenvolve em regiões tropicais e temperadas, sendo originária da África Ocidental, do Sudeste Asiático e da América tropical. No Brasil, existem cerca de 140 espécies de dioscoreáceas que crescem em áreas de Mata Atlântica, Cerrado, Pampa, Caatinga e Amazônia. As espécies *D. alata* e *D. cayennensis-rotundata* estão entre as dioscoreáceas mais cultivadas no mundo. A cultura do inhame tem grande importância econômica, alimentar e social, principalmente na região do Nordeste do Brasil. Em 2019, o Brasil ocupou o décimo lugar na produção mundial de inhame e a região Nordeste (Pernambuco e Paraíba) foram os maiores produtores deste cultivo. Por ser uma cultura de propagação vegetativa através de túberas-semente, o inhame pode acumular uma série de fitopatógenos, dentre eles, os vírus. Até o presente momento, as espécies virais relatadas no inhame pertencem aos gêneros: *Potyvirus*, *Macluravirus*, *Badnavirus*, *Ampelovirus*, *Cucumovirus*, *Potexvirus*, *Aureusvirus*, *Sadwavirus* e *Carlavirus*. Dentre esses, os potyvírus (família *Potyviridae*) são os que induzem os maiores danos econômicos na cultura. Estes vírus são monopartidos, possuem partículas flexuosas com genoma viral de RNA de fita simples de senso positivo entre 9,4 - 11,0 kb de comprimento e são transmitidos por pulgões de forma não-persistente. Entre os potyvírus, o yam mosaic virus (YMV) e o yam mild mosaic virus (YMMV) são os principais agentes virais em todo o mundo. O presente estudo visou o diagnóstico de infecções virais em espécies de dioscoreáceas de áreas produtoras do Nordeste do Brasil usando o sequenciamento de alto rendimento (*High-Throughput Sequencing*, HTS). Para tanto, foram coletadas plantas de *D. cayennensis-rotundata*, *D. alata*, *D. bulbifera*, *D. trifida*, *D. altissima* e *D. subhastata*. O RNA total extraído foi submetido ao sequenciamento Illumina visando uma análise global e montagem dos genomas (quase)completos virais. Os *contigs* foram montados com SPAdes v. 3.9 e submetidos à análise tBlastX usando um banco de dados de sequências de genoma virais (refSeq março de 2020) implementado no software Geneious v 11.0 (R11). Sequências genômicas virais (quase) completas virais foram montadas de amostras de *D. cayennensis-rotundata*, e classificadas pertencendo às espécies *Yam mosaic virus* (YMV, *Potyvirus*), *Dioscorea bacilliform AL virus* e *Dioscorea bacilliform RT virus* (*Badnavirus*), e *Dioscorea mosaic-associated virus* (*Sadwavirus*). A partir de RNA total de *D. alata* foram montadas sequências genômicas das espécies *Yam mild mosaic virus* (YMMV, *Potyvirus*), *Dioscorea bacilliform AL virus* (*Badnavirus*) e *dioscorea virus A* (Família *Betaflexiviridae*). Nenhum vírus foi detectado nas outras espécies de *Dioscorea*. O YMV foi detectado por ELISA apenas em plantas de *D. cayennensis-rotundata* e em todos os locais de amostragem (Alagoas, Bahia, Maranhão, Paraíba e Pernambuco), enquanto o YMMV foi encontrado apenas em *D. alata*. A caracterização de três sequências genômicas de YMV foi realizada e uma foi validada com o sequenciamento de Sanger. Análises espaço-temporais foram realizadas com 26 sequências de nucleotídeos (nt) do gene codificador da proteica (CP) de YMV com o intuito de estimar a taxa evolutiva e o tempo do ancestral comum mais recente (*time to the most recent common ancestral*, TMRCA). A análise temporal pelo programa BEAST revelou como melhor modelo de relógio o relaxado lognormal não correlacionado (UCLN), resultando numa taxa média de evolução de $1,76 \times 10^{-3}$ (95% HDP: 2.45×10^{-4} - 3.90×10^{-3}) substituições por sítio por ano, valores estimados médios de TMRCA em 168,68 anos, (95% HPD: 48.56-363.28) e a maior probabilidade de origem geográfica de YMV na Nigéria. Segundo a inferência Bayesiana, as quatro sequências de YMV montadas neste estudo foram agrupadas em um novo grupo monofilético (grupo X). Além disso, foi caracterizada pela primeira vez a sequência genômica (quase)completa de YMMV proveniente de *D. alata* no Brasil. Os resultados das múltiplas comparações par a par de sequências utilizando sequências genômicas completas revelaram uma baixa identidade de 87% com a sequência genômica do isolado brasileiro de *D. trifida*. De acordo com as análises filogenéticas, baseadas nas sequências de nucleotídeos da CP de YMMV, propomos a reestruturação dos grupos de YMMV

em dez filogrupos (I-X). As sequências de YMMV deste estudo foram agrupadas em um novo grupo proposto (IX), constituído por isolados de *D. alata* do Togo, Nigéria e Brasil. Por fim, os dados apresentados neste estudo contribuem para revelar aspectos-chave da história epidemiológica dos potyvírus que infectam o inhame no Brasil.

Palavras-chave: *Yam mosaic virus*, *Yam mild mosaic virus*, *Dioscorea* spp., HTS, Estudos filogeográficos e geoespaciais de YMV.

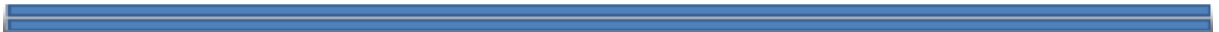
GENERAL ABSTRACT

Yam (*Dioscorea* L., Dioscoreaceae Family) grows in tropical and temperate regions around the world and is native to West Africa, Southeast Asia and tropical America. In Brazil, there are about 140 species of Dioscoreaceae that grow in the Atlantic Forest, Cerrado, Pampa, Caatinga and Amazon areas. *D. alata* and *D. cayennensis-rotundata* are among the most cultivated species in the world. The yam crop has economic, food and social importance, mainly in the Northeast of Brazil. In 2019, Brazil ranked tenth in the world production of yam and the Northeast region (Pernambuco and Paraíba) were the largest producers of this crop. As it is a vegetatively propagated culture through seed tubers, it accumulates several pathogens, including viruses. So far, the viral species reported in yam belong to the genera: *Potyvirus*, *Macluravirus*, *Badnavirus*, *Ampelovirus*, *Cucumovirus*, *Potexvirus*, *Aureusvirus*, *Sadwavirus* and *Carlavirus*. However, potyviruses of the *Potyviridae* family are considered to cause the greatest economic damage to the crop. These viruses are monopartite, have flexuous particles with a positive-sense single-stranded RNA viral genome of 9.4 - 11.0 kb in length and are transmitted by aphids in a non-persistent manner. Among the potyviruses, yam mosaic virus (YMV) and yam mild mosaic virus (YMMV) are the main viral agents worldwide. For this reason, in this study, we performed a diagnosis of viral infections in *Dioscorea* species from the producing areas of Northeast Brazil using High-Throughput Sequencing (HTS). To this extent, plants of *D. cayennensis-rotundata*, *D. alata*, *D. bulbifera*, *D. trifida*, *D. altissima* and *D. subhastata* were collected, and the total RNA extracted was submitted to Illumina sequencing for global analysis and assembly of the complete viral genomes. The obtained contigs were assembled with SPAdes v. 3.9 and subjected to tBlastX analysis using a viral genomic sequence database (refSeq March 2020) implemented in Geneious v 11.0 (R11) software. Nearly-complete viral genomic sequences were assembled from samples of *D. cayennensis-rotundata* and classified as belonging to the species *Yam mosaic virus* (*Potyvirus*), *Dioscorea bacilliform AL virus* and *Dioscorea bacilliform RT virus* (*Badnavirus*) and *Dioscorea mosaic-associated virus* (*Sadwavirus*). From the total RNA of *D. alata*, genomic sequences of the species *Yam mild mosaic virus* (*Potyvirus*), *Dioscorea bacilliform AL virus* (*Badnavirus*) and *dioscorea virus A* (Family *Betaflexiviridae*) were assembled. No virus was detected in the other *Dioscorea* species. YMV was detected by ELISA only in *D. cayennensis-rotundata* plants and in all sampling sites (Alagoas, Bahia, Maranhão, Paraíba and Pernambuco), while YMMV was found only in *D. alata*. The characterization of three YMV genomic sequences was performed and one was validated with Sanger sequencing. Spatio-temporal analyses were performed with 26 nucleotides (nt) sequences of the YMV CP gene in order to estimate the evolutionary rate and time of the most recent common ancestor (Time to the Most Recent Common Ancestor, TMRCA). The temporal analysis by BEAST revealed the relaxed uncorrelated lognormal (UCLN) as the best clock model, resulting in an average evolution rate of 1.76×10^{-3} (95% HDP: 2.45×10^{-4} - 3.90×10^{-3}) site substitutions per year, estimated mean time to the most recent common ancestor at 168.68 years ago (95% HPD: 48.56-363.28), and the highest probability of the geographic origin of YMV in Nigeria. According to the Bayesian inference,

the four YMV sequences assembled in this study were grouped into a new monophyletic group (group X). In addition, the near-complete genomic sequence of YMMV from *D. alata* in Brazil was characterized for the first time. Results of multiple comparisons using complete genomic sequences revealed a low identity of 87% with the genomic sequence of the Brazilian isolate of *D. trifida*. According to the phylogenetic analysis, based on the nt sequences of the YMMV CP, we propose the restructuring of the YMMV groups into ten phylogroups (I-X). The YMMV sequences from this study were grouped into a proposed new group (IX) consisting of isolates of *D. alata* from Togo, Nigeria and Brazil. Finally, the data presented in this study contribute to revealing key aspects of the epidemiological history of the potyviruses that infect yams in Brazil.

Keywords: *Yam mosaic virus*, *Yam mild mosaic virus*, *Dioscorea* spp., HTS, Phylogeographic and Geospatial studies of YMV.

CAPÍTULO I



Introdução Geral

ANÁLISE GENÔMICA DE POTYVÍRUS ASSOCIADOS A ESPÉCIES DE DIOSCOREÁCEAS NO BRASIL

INTRODUÇÃO GERAL

1. A cultura do inhame: aspectos nutricionais e econômicos

A família Dioscoreaceae contém mais de 680 espécies de plantas distribuídas em nove gêneros (ZAPPI *et al.*, 2015; COUTO, 2020; WCVP, 2022), as quais se desenvolvem em zonas tropicais e temperadas. As regiões da África Ocidental, Sudeste Asiático e América tropical são consideradas como os três centros de origem (ANDRES; ADEOLUWA; BHULLAR, 2016; PROTA, 2017). A maioria das espécies de dioscoreáceas são conhecidas popularmente ao redor do mundo especialmente como yams, inhame, ñame, igname, niam, enyame, nyami, além de outros nomes comuns (COUTO *et al.*, 2018).

Segundo a Flora do Brasil (2020), nas Américas ocorrem cerca de 25% do total das espécies de dioscoreáceas. No entanto, somente no Brasil existem cerca de 140 espécies de plantas que crescem nas regiões de Mata Atlântica, Cerrado, Pampa, Caatinga e Amazônia (ZAPPI *et al.*, 2015). Na Tabela 1 são apresentadas informações sobre a distribuição geográfica de algumas espécies tais como a *D. cayennensis-rotundata* Lam., *D. alata* L., *D. bulbifera* L., *D. trifida* L.f., *D. altissima* Lam. e *D. subhastata* Vell. (COUTO; FRAGA, 2020).

Tabela 1 - Espécies de dioscoreáceas, nome comum e sua distribuição geográfica no Brasil.

Nome da espécie	Nome comum	Distribuição geográfica	Link de acesso ao táxon
<i>Dioscorea cayennensis-rotundata</i> Lam.	“cará-da-costa”, “inhame”, “inhame-de-pernambuco”	Ocorrências confirmadas Nordeste: Bahia, Ceará, Paraíba, Pernambuco. Sudeste: Minas Gerais, Rio de Janeiro.	http://servicos.jbrj.gov.br/flora/search/Dioscorea_cayennensis
<i>Dioscorea alata</i> L.	“cará-inhame”, “cará”, “inhame”	Ocorrências confirmadas Norte: Amazonas, Rondônia, Roraima. Nordeste: Maranhão. Centro-Oeste: Distrito Federal, Mato Grosso. Sudeste: Minas Gerais, Rio de Janeiro, São Paulo. Sul: Paraná. Possíveis ocorrências Norte: Acre, Amapá. Nordeste: Alagoas, Ceará.	http://servicos.jbrj.gov.br/flora/search/Dioscorea_alata

<i>Dioscorea bulbifera</i> L.	“cará-de-árvore”, “cará-do-ar”, “cará-moela”	Ocorrências confirmadas Norte: Amazonas. Nordeste: Bahia, Ceará, Maranhão e Pernambuco. Centro-Oeste: Mato Grosso do Sul. Sudeste: Minas Gerais, Rio de Janeiro, São Paulo. Sul: Paraná.	http://servicos.jbrj.gov.br/flora/search/Dioscorea_bulbifera
<i>Dioscorea trifida</i> L.f.	“cará-branco”, “cará-roxo”	Ocorrências confirmadas Norte: Acre, Amazonas, Amapá, Pará, Rondônia, Tocantins. Nordeste: Maranhão, Paraíba, Pernambuco. Centro-Oeste: Distrito Federal, Goiás, Mato Grosso do Sul, Mato Grosso. Sudeste: Minas Gerais. Possíveis ocorrências Norte: Roraima.	http://servicos.jbrj.gov.br/flora/search/Dioscorea_trifida
<i>Dioscorea altissima</i> Lam.	-	amplamente distribuído no Brasil	https://www.gbif.org/es/species/2754951
<i>Dioscorea subhastata</i> Vell.	“cará-redondo-chato”	Ocorrências confirmadas Nordeste: Alagoas, Bahia, Ceará, Paraíba, Pernambuco e Sergipe. Sudeste: Espírito Santo, Minas Gerais, Rio de Janeiro, São Paulo. Sul: Paraná, Rio Grande do Sul, Santa Catarina.	http://servicos.jbrj.gov.br/flora/search/Dioscorea_subhastata

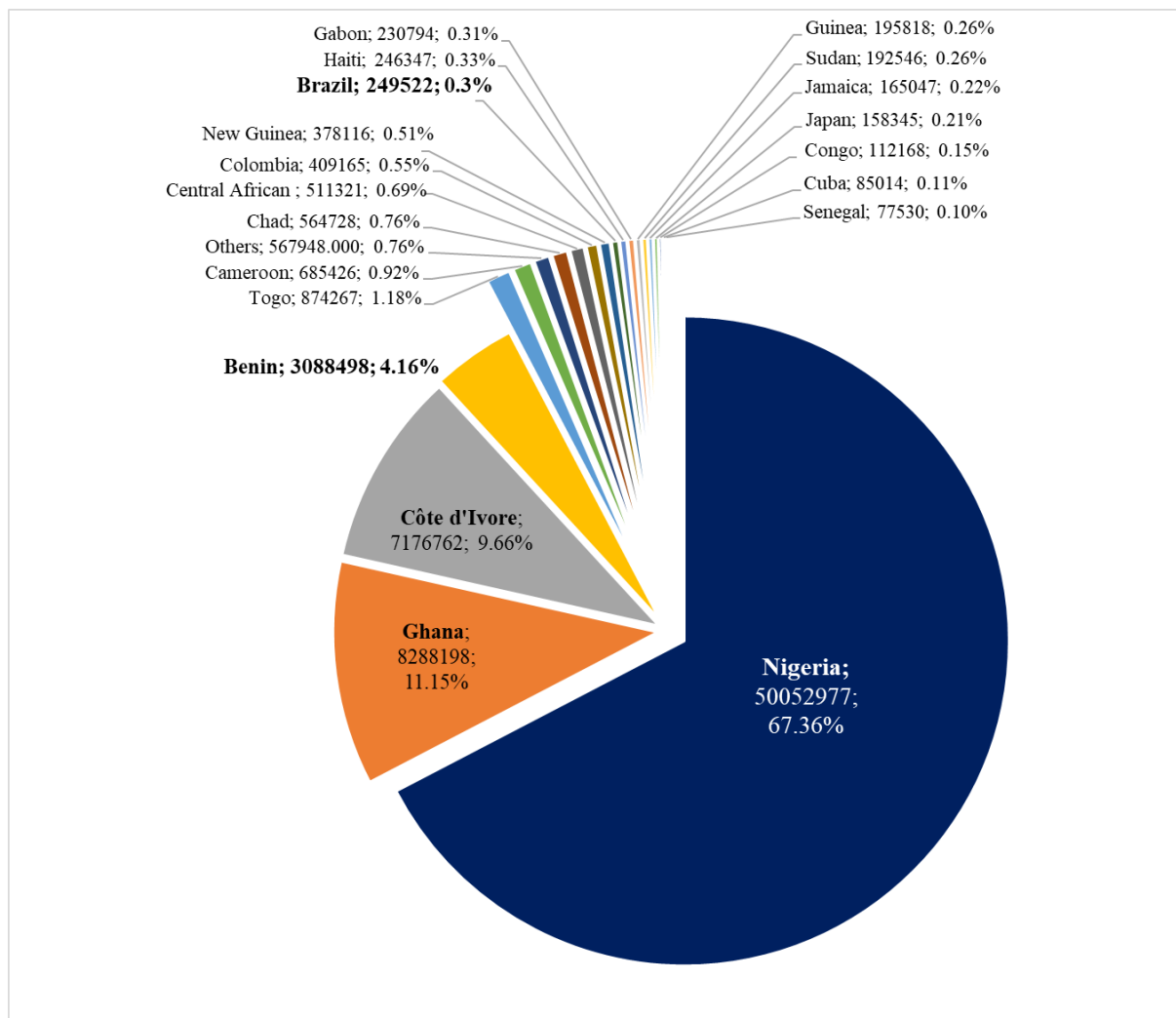
As principais espécies de dioscoreáceas comestíveis que são consideradas fontes básicas de alimento no mundo são: *D. cayennensis* Lam., *D. rotundata* Poir., *D. alata* L., *D. bulbifera* L., *D. esculenta* (Lour.) Burkill, *D. trifida* L., *D. opposita* Thunb., *D. nummularia* Lam., *D. transversa* R. Br., *D. persimilis* Prain. et Burk, e *D. pentaphylla* L. (MALAPA *et al.*, 2005). Cada espécie é nomeada de forma diferente de acordo com o local onde é cultivada. Por exemplo, os nomes comuns de *D. alata* mais utilizados a nível mundial são: “purple yam”, “greater yam Guyana”, “water yam”, “Winged yam”, “Raja ala” (Sri Lanka) e “Ube” (Filipinas) (CHANDRASEKARA; JOSHEPH KUMAR, 2016).

O inhame pode ser consumido cru, cozido, em pó ou sob forma de farinha. As túberas têm vários componentes bioativos tais como mucina, dioscina, dioscorina, alantoína, colina, polifenóis e diosgenina (CHOUDHARY *et al.*, 2012; SHAJEELA *et al.*, 2011), e contêm grandes quantidades de proteínas quando comparadas com outras raízes e tubérculos. As variedades amarelas de inhame contêm caroteno, além de quantidades expressivas de fibra dietética (CHANDRASEKARA; JOSHEPH KUMAR, 2016). A mucilagem das túberas contém glicoproteínas solúveis e fibra alimentar, e vários estudos mostraram que o extrato do inhame possui atividade hipoglicêmica, antimicrobiana e antioxidante (CHAN *et al.*, 2004;

KELMANSON; JA; STADEN, 2000). Além disso, o consumo do inhame pode estimular a proliferação de células epiteliais gástricas e aumentar as atividades de enzimas digestivas no intestino delgado (CHEN *et al.*, 2003). Nutricionalmente, estas culturas têm um grande potencial para servir como fonte de energia dietética, na forma de carboidratos. A relativa importância da cultura de raízes e túberas se evidencia na produção global de aproximadamente 836 milhões de toneladas somente no ano de 2019 (FAOSTAT, 2021).

Segundo dados de 2019, a Nigéria, Gana, Costa do Marfim e Benin são os maiores produtores de inhame, responsáveis por aproximadamente 93,3% da produção mundial (~68,6 milhões de toneladas/ano), em uma área plantada de quase 8,2 milhões de hectares. Somente a Nigéria produziu acima de 50 milhões de toneladas/ano com uma produtividade média de 8.016 kg/ha, e, portanto, gerou 67,36% da produção mundial. No entanto, países como o Guayana, Japão e a ilha do Caribe Santa Lúcia, que dispõem de maior nível tecnológico, chegam a alcançar uma produtividade superior a 22.000 kg/ha (FAOSTAT, 2021). No ano de 2019, o Brasil ficou em décimo lugar na produção mundial de inhame, com 249.522 ton, o que representou 0,34% da produção mundial (Figura 1). De acordo com o censo agropecuário do Brasil de 2017, as regiões Nordeste (Pernambuco e Paraíba), Sudeste (Espírito Santo e Minas Gerais) e o Norte (Rondônia) foram os maiores produtores de cará, e somente a região Nordeste é responsável por 50% da produção nacional.

Figura 1 - Produção (ton) mundial de inhame de 2019. O gráfico representa a distribuição por país da produção em toneladas (ton), e expressos em porcentagem (%).



Fonte: Elaborado com dados de FAOSTAT (2021), (<http://www.fao.org/faostat/en/#data/QC>)

2. Viroses na cultura do inhame

Dentre os principais grupos de vírus relatados que infectam a cultura do inhame (Tabela 2), destacam-se os pertencentes aos gêneros *Potyvirus* (família *Potyviridae*) (ALEMAN *et al.*, 1996; BOUSALEM; DOUZERY; FARGETTE, 2000; BOUSALEM *et al.*, 2003; RABELO-FILHO *et al.*, 2013; FUJI; NAKAMAE, 1999), *Macluravirus* (*Potyviridae*) (FILLOUX *et al.*, 2015; KONDO; FUJITA, 2012; LAN *et al.*, 2018; HU *et al.*, 2018) *Badnavirus* (*Caulimoviridae*) (BÖMER *et al.*, 2016; BOUSALEM *et al.*, 2009; BRIDDON *et al.*, 1999; SEAL *et al.*, 2014; SUKAL *et al.*, 2017; UMBER *et al.*, 2017), *Ampelovirus* (*Closteroviridae*) (DEY *et al.*, 2019; MARAIS *et al.*, 2020), *Cucumovirus* (*Bromoviridae*) (ENI *et al.*, 2008, 2013), *Potexvirus* (*Alphaflexiviridae*) (AMUSA *et al.*, 2003; ANZALA *et al.*, [s.d.]; CHEN *et al.*, 2002; MAMBOLE *et al.*, 2014), *Aureusvirus* (*Tombusviridae*) (MENZEL; THOTTAPPILLY;

WINTER, 2014), *Sadwavirus* (Secoviridae) (HAYASHI *et al.*, 2017), e *Carlavirus* (Betaflexiviridae) (SILVA *et al.*, 2019).

Tabela 2. Lista de espécies virais reportadas infectando espécies de dioscoreáceas no Brasil e no mundo.

Família	Gênero	Espécie	Número de acesso no GenBank de genomas completos	Referência
<i>Alphaflexiviridae</i>	<i>Potexvirus</i>	<i>Yam virus X</i> (YVX)	KJ711908	(MAMBOLE <i>et al.</i> , 2014)
		yam potexvirus 1	Não determinado	
<i>Betaflexiviridae</i>	Unclassified	<i>Yam virus Y</i> (YVY)	MK782911	(SILVA <i>et al.</i> , 2019)
	<i>Foveavirus</i>	dioscorea virus A	LC467961	(HAYASHI <i>et al.</i> , 2016)
	<i>Carlavirus</i>	yam latent virus	KJ789130	ICTV, NCBI (Não publicado)
<i>Bromoviridae</i>	<i>Cucumovirus</i>	<i>Cucumber mosaic virus</i> (CMV)	Não determinado	(ENI <i>et al.</i> , 2013)
<i>Caulimoviridae</i>	<i>Badnavirus</i>	<i>Dioscorea bacilliform AL virus</i> (DBALV)	KX008573	(UMBER <i>et al.</i> , 2017)
		<i>Dioscorea bacilliform AL virus 2</i> (DBALV 2)	KY827395	(SUKAL <i>et al.</i> , 2017)
		<i>Dioscorea bacilliform ES virus</i> (DBESV)	KY827394	(SUKAL <i>et al.</i> , 2017)
		<i>Dioscorea bacilliform RT virus 1</i> (DBRTV1)	NC_038986	(BÖMER <i>et al.</i> , 2016)
		<i>Dioscorea bacilliform RT virus 2</i> (DBRTV2)	KX008577	(BÖMER <i>et al.</i> , 2016)
		<i>Dioscorea bacilliform RT virus 3</i> (DBRTV3)	MF476845	(BÖMER <i>et al.</i> , 2018)
		<i>Dioscorea bacilliform SN virus</i>	DQ822073	(SEAL; MULLER, 2007)
		<i>Dioscorea bacilliform TR virus</i> (DBTRV)	KX430257	(UMBER <i>et al.</i> , 2017)
		<i>Dioscorea bacilliform virus</i> (DBV)	NC_009010	(SEAL; MULLER; 2007)
<i>Closteroviridae</i>	<i>Ampelovirus</i>	yam asymptomatic virus 1 (YaV1)	MT409627	(MARAIS <i>et al.</i> , 2020)
<i>Potyviridae</i>	<i>Macluravirus</i>	dioscorea alata macluravirus	Não determinado	(FILLOUX <i>et al.</i> , 2015)
		dioscorea esculenta macluravirus	Não determinado	(FILLOUX <i>et al.</i> , 2015)
		<i>Yam chlorotic necrotic mosaic virus</i>	NC_038561	(ZHANG, P. <i>et al.</i> , 2016)
	<i>Potyvirus</i>	<i>Yam mild mosaic virus</i> (YMMV)	NC_019412	(RABELO-FILHO <i>et al.</i> , 2013)
		<i>Yam mosaic virus</i> (YMV)	NC_004752	(ALEMAN <i>et al.</i> , 1996)
<i>Secoviridae</i>	<i>Sadwavirus</i>	<i>Dioscorea mosaic-associated virus</i> (DMaV)	KU215538, KU215539	(HAYASHI <i>et al.</i> , 2017)
	<i>Fabavirus</i>	<i>Broad bean wilt virus 2</i> (BBWV-2)	Não determinado	(HE <i>et al.</i> , 2021)
<i>Tombusviridae</i>	<i>Aureusvirus</i>	<i>Yam spherical virus</i> (YSV)	KF482072	(MENZEL; THOTTAPPILLY; WINTER, 2014)

No entanto, a cultura ainda é considerada negligenciada e os estudos de diversidade viral no inhame são incipientes nas condições brasileiras (SIQUEIRA, 2011). Em escala mundial, o impacto de infecções virais no rendimento e na qualidade das túberas podem levar a 100% de perda de produção. Em um trabalho de pesquisa realizada no Instituto Internacional de Agricultura Tropical (IITA) em Ibadan, Nigéria, Adeniji et al. (2012) mensuraram perdas de rendimento de 52 a 65% de *D. rotundata* infectadas pelo vírus do mosaico do inhame (YMV, gênero *Potyvirus*). O YMV é o vírus mais prevalente em *D. cayennensis-rotundata* (ASALA et al., 2012) e ocasiona sintomas de mosaico, manchas, faixas verdes nas nervuras (THOUVENEL; FAUQUET, 1679), bem como deformação foliar e crescimento atrofiado levando à redução da produção das sementes-túberas (BÖMER et al., 2019). Além disso, o vírus do mosaico suave do inhame (YMMV, gênero *Potyvirus*) e os vírus baciliformes das dioscoreáceas (DBVs, gênero *Badnavirus*) estão disseminados nas principais áreas produtoras de inhame na África Ocidental (AMUSA et al., 2003; KENYON et al., 2001; BOUSALEM et al., 2009). Asala e Alegbejo (2016) demonstraram que o YMV, YMMV e DBV causam doenças importantes em campos de inhame em infecções simples ou mistas. Azeteh et al. (2019) realizaram o monitoramento de campos de produção de inhame na África ocidental (Camarões) para quantificar a incidência de doenças causadas por YMV, YMMV e o vírus do mosaico do pepino (CMV) por RT-PCR. Os autores encontraram que em 75% das amostras avaliadas havia a presença de infecções ocasionadas por YMV e YMMV (AZETEH et al., 2019).

Estudos filogenéticos baseados em sequência de nucleotídeos (nt) da região codificadora da CP de YMMV e YMV suportam uma classificação direta em vários filogrupos. Bousalem et al. (2003) classificaram o YMMV em oito filogrupos (I-VIII) usando 36 sequências virais. Anos mais tarde, Nkere et al. (2020) mostraram a existência de mais três filogrupos (IX, X e XI) utilizando isolados de YMMV a partir de amostras de dioscoreáceas coletadas no Guiné-Bissau, Nigéria, Benin e Camarões entre 2012 e 2013. Já Aleman-Verdaguer et al. (1997) e Bousalem et al. (2000) classificaram os isolados de YMV em nove filogrupos diferentes (I-IX): o grupo I inclui isolados de Burkina Faso; o grupo II de Guadalupe e Costa do Marfim, o grupo III de Benin e Camarões, o grupo IV de Benin, Burkina Faso e Camarões, o grupo V da Guiana Francesa, o grupo VI de Guadalupe e Porto Rico, o grupo VII da Nigéria, o grupo VIII de Guadalupe e o grupo IX de Camarões.

No Brasil, a primeira investigação realizada para inspecionar a presença de um potyvírus em um banco de germoplasma de inhame da Bahia utilizando microscopia eletrônica foi conduzida por de Ávila et al. (1982 apud KITAJIMA, 2020, p. 54). Mais tarde, alguns estudos foram realizados para verificar a possível infecção por YMV em campos dos estados de Sergipe

(BOARI *et al.*, 2005 citado por KITAJIMA, 2020, p. 54), Pernambuco (PIO-RIBEIRO *et al.*, 2006), Paraíba (ANDRADE *et al.*, 2007) e Bahia (COSTA *et al.*, 2013). Pio-Ribeiro *et al.* (2006) analisaram amostras de *D. rotundata* proveniente dos estados da Paraíba e Pernambuco por serologia com DAC-ELISA (*Direct Antigen Coating-ELISA*) contra CMV, TAS-ELISA (*Triple Antibody Sandwich-ELISA*) contra YMV, e por testes moleculares do tipo DNAC-RT-PCR (*direct nucleic acid coating one-step RT-PCR*) usando oligonucleotídeos (*primers*) específicos para YMV e YMMV, e do tipo DNAC-PCR com *primers* degenerados para badnavírus. Como resultado, os autores encontraram reação positiva apenas para o YMV no TAS-ELISA, e por testes moleculares constataram infecções simples de YMV e mistas entre YMV e badnavírus. O primeiro genoma completo do potyvírus YMMV foi caracterizado no ano de 2009 em isolado de plantas de *D. trifida* coletadas no estado de Pernambuco (RABELO-FILHO *et al.*, 2013).

3. Família *Potyviridae*

A família *Potyviridae* é uma das maiores famílias de vírus de plantas constituída atualmente por 12 gêneros (*Arepavirus*, *Bevemovirus*, *Brambyvirus*, *Bymovirus*, *Celavirus*, *Ipomovirus*, *Macluravirus*, *Poacevirus*, *Potyvirus*, *Roymovirus*, *Rymovirus* e *Tritimovirus*) (YANG *et al.*, 2019; SEO *et al.*, 2017; SUSAIMUTHU *et al.*, 2008; CHEN *et al.*, 1999; ROSE *et al.*, 2019; JANSSEN *et al.*, 2005; HU *et al.*, 2018; FELLERS *et al.*, 2009; JOVER-GIL *et al.*, 2018; JO; CHOI; CHO, 2015; STENGER; FRENCH, 2004). Existem ainda mais três espécies que não foram classificadas: *Common reed chlorotic stripe virus* (YUAN *et al.*, 2017), *Longan witches broom-associated virus* (SEO *et al.*, 2017a), e *Spartina mottle virus* (ROSE *et al.*, 2020). Os vírus que pertencem à família *Potyviridae* têm como característica partículas não envelopadas, flexuosas, de 680-900 nm de comprimento e 11-20 nm de diâmetro. O genoma é composto por um RNA de fita simples senso-positivo de 8-11,3 kb (kilo/quilo base) de tamanho, geralmente monopartido, com exceção das espécies pertencentes ao gênero *Bymovirus* que possuem genoma bipartido (PEERENBOOM *et al.*, 1997; YOU; SHIRAKO, 2010; VALLI; GARCÍA; LÓPEZ-MOYA, 2015). Os vírions são constituídos por uma proteína capsidial (CP) de 30 a 47 kDa (TORRANCE *et al.*, 2006) e o RNA, sendo este flanqueado a uma região não traduzida 5'-UTR (5'-*untranslated region*) com uma proteína terminal ligada covalentemente (VPg) e a uma região 3'-UTR seguida de uma porção poliadenilada (cauda poli-A) (VALLI; GARCÍA; LÓPEZ-MOYA, 2015; WYLIE *et al.*, 2017). Segundo o ICTV (Comitê Internacional de Taxonomia de Vírus, *International Committee on Taxonomy of Viruses*) a família é constituída por um total de 228 espécies (ICTV, 2021), que infectam uma ampla gama de hospedeiros, tanto plantas monocotiledôneas como dicotiledôneas (ADAMS, M. J.; ANTONIW; FAUQUET, C. M., 2005; WYLIE *et al.*, 2017), abrangendo cerca de 57 famílias de plantas,

principalmente as da divisão Magnoliophyta (angiospermas ou plantas com flores). No entanto, as diferentes espécies de potyvírus geralmente têm uma gama estreita de hospedeiros, com algumas exceções (BRUNT *et al.*, 1996).

Os membros da família *Potyviridae* com genomas monopartidos (ex. potyvírus) são organizados da seguinte forma: uma longa ORF (fase/quadro de leitura aberta, *Open Reading Frame*) que codifica uma poliproteína, que é clivada por três proteases auto codificadas em proteínas funcionais em regiões semi-conservadas no genoma. Já os membros com genomas bipartidos (ex. *Bymovirus*) codificam duas poliproteínas que são clivadas por duas proteases (ADAMS; ANTONIW; BEAUDOIN, 2005). As regiões amino, central e carboxi-terminal da poliproteína em potyvírus têm uma organização conservada e codificam as proteínas maduras P1 (*protein 1 protease*), HC-Pro (*helper component protease*), P3 (*protein 3*), 6K1 (*six kilodalton protein 1*), CI (*cylindrical inclusion*), 6K2 (*six kilodalton protein 2*), VPg (*viral protein genome-linked*), NIaPro (*nuclear inclusion A*), NIb (*nuclear inclusion B*), e CP (*capsid protein*), ocorrendo o mesmo com a poliproteína codificada pelo RNA1 dos bymovírus, sendo essa poliproteína processada pela protease NIaPro (ADAMS; ANTONIW; BEAUDOIN, 2005). Existe também na maioria dos genomas dos potyvírus uma segunda ORF curta (*Pretty Interesting Potyvirus ORF*, PIPO), que foi identificada dentro da região codificadora de P3, onde transcritos de PIPO são expressos a partir de um mecanismo de deslizamento da polimerase na sequência conservada de GA₆, onde a RNA polimerase lê a mesma base mais de uma vez, criando inserções de bases, gerando diferentes proteínas (CHUNG *et al.*, 2008; OLSPERT *et al.*, 2015; RODAMILANS *et al.*, 2015).

Em todos os membros da família *Potyviridae*, a ordem das regiões de codificação das proteínas virais é geralmente conservada. No entanto, em alguns vírus (ex. cassava brown streak viruses, CBSV, gênero *Ipomovirus*) podem existir uma sequência genômica adicional entre a NIb e a CP que codifica para a proteína “Ham1-like” (atividade de nucleosídeo-trifosfato difosfatase) (WINTER *et al.*, 2010). Por sua vez, a região de codificação da P1 pode estar ausente em outros vírus (ex. gêneros *Macluravirus* e *Bevemovirus*) (KONDO; FUJITA, 2012; SEO *et al.*, 2017b), assim como a HC-Pro (ROSE *et al.*, 2019; YOU; SHIRAKO, 2010) e/ou outras regiões (SUSAIMUTHU *et al.*, 2008). Outras características típicas dos vírus da família *Potyviridae* são: (1) a replicação é citoplasmática, sendo iniciada em um complexo de replicação viral nas vesículas membranosas (induzidas pela ação da proteína 6K2) dos locais de saída do retículo endoplasmático (ER); (2) a tradução das proteínas é realizada diretamente do RNA genômico; (3) a infecção é restrita apenas à plantas e (4) a maioria dos membros desta família é transmitida por artrópodes, com exceção dos membros do gênero *Bymovirus*, que são

transmitidos por fungos plasmodióforos (ADAMS; SWABY; JONES, 1988; PEERENBOOM *et al.*, 1997). Algumas das principais características dos vírus dos diferentes gêneros da família *Potyviridae* e do gênero *Potyvirus* estão apresentadas abaixo e mais detalhadamente descritas no **item 3.1.**

Os membros do gênero ***Bevemovirus*** não apresentam a região codificadora da proteína P1. A proteína HC-Pro não apresenta os motivos conservados de transmissão por pulgões (R/KITC e PTK), que leva a dedução que esses vírus não são transmitidos por pulgões. A única espécie reconhecida pelo ICTV é o *Bellflower veinal mottle virus*, na qual a sequência genômica compartilha uma identidade de menos de 27% com sequências dos macluravírus (SEO *et al.*, 2017). O gênero *Brambyvirus*, assim como observado no gênero *Bevemovirus*, é atualmente constituído por apenas uma única espécie (*Blackberry virus Y*, BVY) encontrada infectando espécies de *Rubus*. Difere-se de outros gêneros por codificar uma P1 muito grande (83,6 kDa) contendo um domínio AlkB (proteínas de reparo de ácido nucléico) (SUSAIMUTHU *et al.*, 2008). Os ortólogos AlkB são encontrados em procariotos, eucariotos e vírus, e são envolvidos no reparo de ácidos nucleicos alquilados (AAS *et al.*, 2003). Também foi especulado que esses ortólogos podem estar envolvidos na supressão do silenciamento de genes (ARAVIND; KOONIN, 2001). O vírion monopartido é constituído por genoma de ssRNA (fita simples de RNA, *single-stranded RNA*) de aproximadamente 11 kb e possui organização genômica idêntica à de outros membros monopartidos da família *Potyviridae*. Se desconhece a expressão de sintomas em plantas herbáceas, assim como não se conhece o vetor de transmissão (SUSAIMUTHU *et al.*, 2008).

O gênero ***Bymovirus***, como mencionado anteriormente, possui um genoma dividido (bipartido). Os vírions medem entre 500-600 nm, e contêm um RNA (RNA1) de 7,5 a 8 kb (250-300 nm de comprimento). O RNA2 possui 3,5 a 4 kb em tamanho, e possui baixa identidade nucleotídica com a sequência do RNA 1, exceto para a região extremidade 5'. Os RNA1 e RNA2 codificam para dois distintos polipeptídeos que sofrem atividade proteolítica para dar origem às proteínas virais funcionais. O RNA1 difere por não possuir a P1 e a HC-Pro. O RNA2 codifica a P2-1, que possui domínios com similaridade de sequência com a proteína HC-Pro de potyvírus (CHEN *et al.*, 1999; REVERS; GARCÍA, 2015; VALLI; GARCÍA; LÓPEZ-MOYA, 2015; YOU; SHIRAKO, 2010) e a proteína P2-2 que tem alto peso molecular e acredita-se desempenhar um papel na transmissão do vetor (YOU; SHIRAKO, 2010). Os membros deste gênero são transmitidos de forma persistente pelo protista *Polymyxa graminis* (Plasmodiophorales), que é um parasita de raiz capaz de permanecer infectivo em esporos de repouso viáveis (ADAMS; SWABY; JONES, 1988; JIANG *et al.*, 2020; PEERENBOOM *et al.*,

1997). Cada membro desse gênero tem uma gama restrita de hospedeiros e infecta plantas da família Gramineae (WYLIE *et al.*, 2017).

Já o gênero ***Ipomovirus*** se diferencia dos demais membros da família *Potyviridae* pelo seu modo de transmissão (não-persistente) e pelo vetor (mosca-branca) (DOMBROVSKY; REINGOLD; ANTIGNUS, 2014; WYLIE *et al.*, 2017). O vírion contém uma única molécula de ssRNA de aproximadamente 9,7 kb (JANSSEN *et al.*, 2005; LI *et al.*, 2008; MBANZIBWA *et al.*, 2011), cuja organização genômica é semelhante ao dos potyvírus (*Sweet potato mild mottle virus*, SPMMV) com uma P1 de alto peso molecular (~83 kDa), e diferentemente dos bromovírus com ausência do domínio AlkB. Os genomas de algumas espécies de ipomovírus (*Cucumber vein yellowing virus*, CVYV; *Squash vein yellowing virus*, SqVYV) diferenciam-se pela codificação de duas proteases serina, P1a e P1b, mas não de HC-Pro (LI *et al.*, 2008), e em outro caso pela codificação de apenas uma protease do tipo serina, P1b, e também sem a HC-Pro (*Cassava brown streak virus*, CBSV) (LI *et al.*, 2008). Espécies pertencentes a esse gênero infectam uma ampla gama de hospedeiros (WINTER *et al.*, 2010).

Os membros do gênero ***Macluravirus*** são transmitidos por pulgões de forma não-persistente. No entanto como os potyvírus, possuem uma molécula de ssRNA de cerca de 8,0 kb. A identidade das sequências de aminoácidos da CP entre os macluravírus e os potyvírus é de 14 a 23% (WYLIE *et al.*, 2017). Esses vírus podem infectar espécies de plantas de até nove famílias. Atualmente dez espécies de macluravírus são reconhecidas, incluindo a *Oxyphylla mosaic virus* (AloMV), *Artichoke latent virus* (ArLV) e *Chinese yam necrotic mosaic virus* (CYNMV). Análises de comparação dos genomas revelam oito sítios de clivagem conservados reconhecidos pela HC-Pro ou NIa-Pro, que clivam a poliproteína em nove proteínas maduras (HC-Pro, P3, 7K, CI, 9K, VPg, NIa, NIb e CP). O gênero *Macluravirus* caracteriza-se também por membros que não possuem a região que codifica a P1. Além disso, o cistron HC-Pro é menor do que o encontrado em sequências genômicas dos membros do gênero *Potyvirus* (HU *et al.*, 2018; KONDO; FUJITA, 2012; MINUTILLO *et al.*, 2015).

A organização genômica de membros que pertencem ao gênero ***Poacevirus*** é similar à de membros dos gêneros *Potyvirus*, *Rymovirus*, *Roymovirus* e *Tritimovirus*. Três espécies são reconhecidas pelo ICTV: *Triticum mosaic virus* (TriMV), *Caladenia virus A* (CalVA) e *Sugarcane streak mosaic virus* (SCSMV). O TriMV é transmitido por ácaros, da mesma forma que os tritimovírus e, além disto, os poacevírus e o tritimovírus são filogeneticamente mais próximos (FELLERS *et al.*, 2009). Os vetores de CalVA e SCSMV ainda são desconhecidos (WYLIE *et al.*, 2012; XU *et al.*, 2010). Outra característica do TriMV é que o RNA viral tem uma região 5'-UTR muito longa em comparação ao CalVA e ao SCSMV (TATINENI *et al.*,

2012). Os vírions dos poacevírus contêm genomas de fita simples de RNA de 9,7-10,2 kb, e os motivos envolvidos na transmissão, KITC, PTK (HC-Pro) e DAG (CP), podem estar ausentes em alguns vírus (ex. SCSMV) (XU *et al.*, 2010). Além disso, a P1 do TriMV funciona como supressor viral invés da HC-Pro (TATINENI *et al.*, 2012). Os vírus TriMV e SCSMV infectam gramíneas (SEIFERS; MARTIN; FELLERS, 2010), enquanto o CalVA infecta orquídeas (WYLIE *et al.*, 2012).

As espécies *Rose yellow mosaic virus* (RYMV) e *Passiflora edulis symptomless virus* (PeSV) pertencem ao **gênero *Rymovirus***. Em relação aos motivos conservados na proteína HC-Pro, tanto KITC quanto PTK não foram encontrados. No entanto, um presumível motivo (C-2X-C) foi identificado na região N-terminal da CP de PeSV e RYMV, estando potencialmente envolvido na transmissão por ácaros eriofídeos (JOVER-GIL *et al.*, 2018; MOLLOV; LOCKHART; ZLESACK, 2013). A identidade entre sequências da poliproteína à nível de aa é baixa, variando de 13% (com sequências de alguns bymovírus) a 23% (com alguns potyvírus). Os genomas de RYMV e PeSV têm 9,5 kb e 9,9 kb, respectivamente, e a poliproteína é processada de forma similar aos potyvírus (JOVER-GIL *et al.*, 2018; MOLLOV; LOCKHART; ZLESACK, 2013; WYLIE *et al.*, 2017). Os tamanhos dos cistrons P3 e CI são menores quando comparados a sequências genômicas de outros membros da família *Potyviridae* e o cistron 6K1 está completamente ausente em RYMV (MOLLOV; LOCKHART; ZLESACK, 2013), mas presente em PeSV (JOVER-GIL *et al.*, 2018).

Atualmente o **gênero *Rymovirus*** inclui três espécies (*Agropyron mosaic virus*-AgMV, *Hordeum mosaic virus*-HoMV e *Ryegrass mosaic virus*-RGMV). Estes vírus são conhecidos por serem transmitidos de forma semi-persistente por espécies de ácaros eriofídeos adaptados ao hospedeiro (WYLIE *et al.*, 2017). Os vírus AgMV (FRENCH; STENGER, 2005) e RGMV (JO; CHOI; CHO, 2015) são transmitidos pelo ácaro da ferrugem dos cereais, *Abacarus hystrix*, mas o vetor de HoMV ainda não é conhecido (FRENCH; STENGER, 2005; WYLIE *et al.*, 2017). Os rymovírus compartilham uma relação monofilogenética com os potyvírus (FRENCH; STENGER, 2005). Os genomas são monopartidos e os vírions contêm moléculas de ssRNA de 9,4-9,5 kb com organização genômica semelhante aos potyvírus. Além disso, possuem uma gama de hospedeiros limitada, e geralmente infectam espécies da família das gramíneas (WYLIE *et al.*, 2017).

Os membros do **gênero *Tritimovirus***, que são atualmente representados por seis espécies virais, são transmitidos de forma semi-persistente por espécies de ácaros erióficos. Os vírions contêm moléculas de ssRNA de 9,2-9,6 kb (GOTZ; MAISS, 1995; STENGER; FRENCH, 2004; WYLIE *et al.*, 2017), com uma estrutura genômica semelhante à dos potyvírus (STENGER *et*

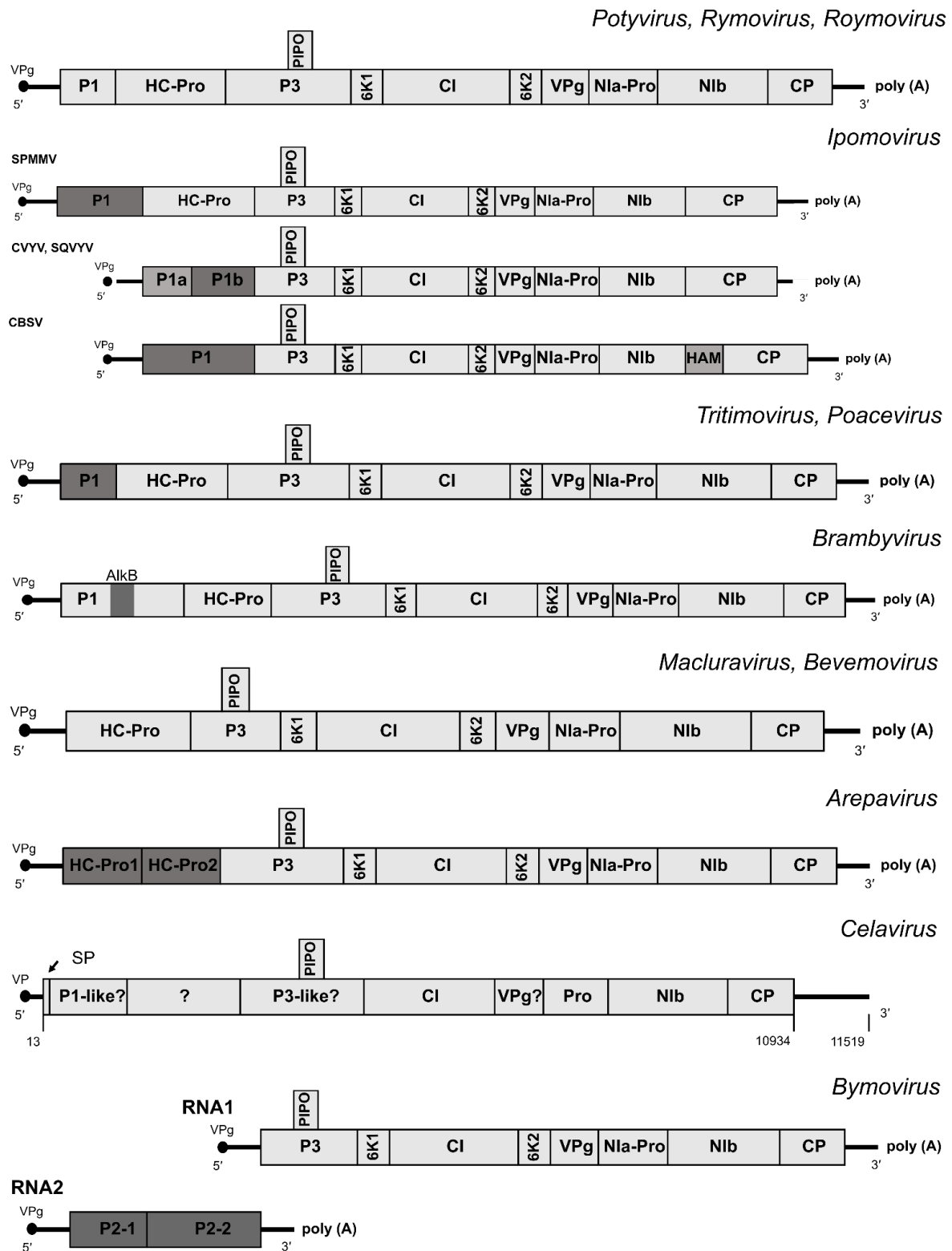
al., 1998). No entanto, os motivos de HC-Pro (do aglomerado de cinco cisteínas, C-8X-C-13X-C-4X-C-2X-C; ou apenas um aglomerado reduzido de V-4X-C-2X-C) e CP (DAG) para transmissão por afídeos estão ausentes em proteínas virais da espécie *Brome streak mosaic virus* (BrSMV) (GOTZ; MAISS, 1995). Segundo Gotz e Maiss (1995) essas variações no genoma explicariam as diferentes relações vetoriais com dos tritimovírus. A sequência da CP de wheat streak mosaic virus (WSMV) apresenta 22-25% de identidade com sequências de alguns membros de potyvírus. Os tritimovírus infectam gramíneas podendo ser associados a uma gama estreita de hospedeiros, como encontrado em experimentos com BrSMV e oat necrotic mottle virus (ONMV) (STENGER; FRENCH, 2004), ou apresentam uma ampla gama como o caso de WSMV. Os vírus deste gênero possuem a proteína P1 em vez de HC-Pro. Essas proteínas P1 atuam como supressores de silenciamento de genes da planta hospedeira (STENGER *et al.*, 1998; VALLI; GARCÍA; LÓPEZ-MOYA, 2015; WYLIE *et al.*, 2017).

Em julho de 2019, em Berlim na Alemanha, o ICTV criou duas novas espécies virais (*Areca palm necrotic ringspot virus* - ANRSV e *Areca palm necrotic spindle-spot virus* - ANSSV) e, portanto, **um novo gênero denominado Arepavirus** pertencente à família *Potyviridae* (WYLIE *et al.*, 2017; YANG *et al.*, 2018, 2019). O ANSSV foi isolado de plantas sintomáticas de *Areca catechu* em 2017 (YANG *et al.*, 2018, 2019). O genoma do arepavírus e seus produtos proteicos mostram a organização genômica típica de um macluravírus. A poliproteína sofre processo proteolítico pelas proteases HC-Pro ou NIa-Pro resultando em nove proteínas maduras (HC-Pro, P3, P3N-PIPO, 6K1, CI, 6K2, VPg, NIa-Pro, NIb e CP), e, portanto, não possui o cistron codificante da P1 (YANG *et al.*, 2018). Interessantemente, trabalhos de anotação do genoma do ANRSV revelaram a existência de duas proteases de cisteína em tandem (HC-Pro1 e HC-Pro2) na extremidade 5' do genoma (WYLIE *et al.*, 2017; YANG *et al.*, 2019). Além disso, os trabalhos de Yang *et al.* (2018, 2019) revelaram que as regiões gênicas que codificam a poliproteína de ANSSV e ANRSV (isolados obtidos de *Areca catechu* L.) compartilham uma identidade de 38-45% a nível de nt e 31-33% de aa com sequências de membros do gênero *Macluravirus*. Os autores demonstraram, portanto, que estas espécies virais formam um gênero diferente dentro da família *Potyviridae*.

Também em julho de 2019 o ICTV criou um **outro gênero (Celavirus)** incluindo uma única espécie (*Celery latent virus*, CeLV) na família *Potyviridae* (ROSE *et al.*, 2019). O CeLV foi descrito pela primeira vez na década de 1970 (BOS; DIAZ-RUIZ; MAAT, 1978; BRANDES; LUISONI, 1966) na Holanda, a partir de plantas assintomáticas de *Apium graveolens* var. *Rapaceum*. Bos e Maat (1978) descreveram que o CeLV infectou experimentalmente 16 hospedeiros incluindo espécies de *Amaranthus*, *Apium*, *Chenopodium*, *Nicotiana*, *Pisum*,

Spinacia e *Trifolium*. Os autores também mostraram a transmissão por semente em aipo, *C. quinoa* e *A. caudatus*. Posteriormente, Rose et al. (2019) também realizaram estudos de transmissão e caracterizaram o genoma completo do CeLV de aipo (*A. graveolens* var. *secalinum*). Os autores observaram que o vírus foi transmitido por seiva e por sementes, e que cinco espécies de pulgões não conseguiram transmitir o CeLV. Nenhuma inclusão cilíndrica citoplasmática foi observada no tecido infectado. Também foram observados vírions filamentosos flexuosos de ~ 900 nm de comprimento que continha um dos maiores genomas já descritos para um potyvírus com cerca de 11,5 kb de tamanho. O RNA viral do CeLV contém uma única ORF flanqueada por uma região não traduzida 5' (5'-UTR) incomumente curta (13 nt) e uma 3'-UTR incomumente longa (586 nt), que não é poliadenilada (ROSE *et al.*, 2019). A PIPO compartilha identidades de sequência à nível de aa muito baixas (~ 18%) com sequências de outros membros da família *Potyviridae* (ROSE *et al.*, 2019). Na poliproteína de celavírus muitos dos motivos característicos de potyvírus não estão presentes, tais como, KITC, IGN, CC / SC, FRNK, PTK e GYCY (HC-Pro) (ROSE *et al.*, 2019). A comparação de sequências da poliproteína de CeLV com outras sequências de membros da família *Potyviridae* é baixa, variando de 15,4 - 44,2% (a nível de nt) e 10 - 14% (a nível de aa). A porcentagem de identidade a nível de nucleotídeos de sequências do gene CP com outras sequências é de 13,8-43,7%. Por último, segundo as análises filogenéticas, o CeLV está localizado na base da árvore em relação a outros potyvírus (CHOI *et al.*, 2022; ROSE *et al.*, 2019). Essas características genômicas incomuns e identidades de sequência abaixo do critério de demarcação de gênero (< 46 %) justificaram a criação de um novo gênero para vírus latentes de aipo, e Rose et al. (2019) propuseram o nome genérico *Celavirus*.

Os esquemas dos mapas genômicos de membros dos doze gêneros da família *Potyviridae* são mostrados na Figura 2 e o resumo das características encontra-se na Tabela 3.

Figura 2 - Mapas genômicos de membros de 12 gêneros da família *Potyviridae*.

A grande ORF (*open reading frame*) é representada por uma caixa dividida por linhas cruzadas em produtos genéticos com seus respectivos nomes dentro de cada subdivisão. A proteína VPg ligada covalentemente é representada por um círculo preto sólido na extremidade 5' e a cauda poli (A) na extremidade 3' indicada por poli (A). Os diferentes recursos são destacados em cinza escuro. Na parte superior direita dos desenhos onde estão indicados os 12 nomes dos gêneros da família *Potyviridae*. Os gêneros são representados por gráfico/s que podem ser compartilhados de acordo com as semelhanças genômicas: (*Potyvirus, Rymovirus, Roymovirus*), *Ipomovirus*, (*Tritimovirus, Poacevirus*); *Brambyvirus*, (*Macluravirus, Bevemovirus*), *Arepavirus, Celavirus* e *Bymovirus*. As

figuras foram adaptadas das seguintes fontes: Potyvirus: yam mosaic virus (YMV) (autor próprio), Rymovirus: ryegrass mosaic virus (RGMV) (JO; CHOI; CHO, 2015), Roymovirus: passiflora edulis symptomless virus (PeSV) e rose yellow mosaic virus (RoYMV) (JOVER-GIL *et al.*, 2018; MOLLOV; LOCKHART; ZLESAK, 2013), Ipomovirus: sweet potato mild mottle virus (SPMMV), cucumber vein yellowing virus (CVYV), squash vein yellowing virus (SqVY) e cassava brown streak virus (CBSV) (WINTER *et al.*, 2010; JANSSEN *et al.*, 2005; LI *et al.*, 2008; MBANZIBWA *et al.*, 2011), Tritimovirus: brome streak mosaic rymovirus (BrSMV) (GOTZ; MAISS, 1995), Poacevirus: triticum mosaic virus (TriMV) (FELLERS *et al.*, 2009), Brambyvirus: blackberry virus Y (BIVY) (SUSAIMUTHU *et al.*, 2008), Macluravirus: yam necrotic mosaic virus (CYNMV) (KONDO; FUJITA, 2012), Bevemovirus: bellflower veinal mottle virus (BVMoV) (SEO *et al.*, 2017b), Arepavirus: areca palm necrotic ringspot disease (ANRSD) (YANG *et al.*, 2019), Celavirus: celery latent virus (CeLV) (ROSE *et al.*, 2019), Bymovirus: barley yellow mosaic virus (BaMYV) (CHEN *et al.*, 1999; YOU; SHIRAKO, 2010).

Tabela 3 - Algumas características de membros de doze gêneros da família *Potyviridae*.

Gênero	No Espécies	Tipo de organização genômica	Tamanho de genoma (kb)	Gama de hospedeiros	Vetor	Características notáveis	Referência
<i>Arepavirus</i>	2	Monopartido	9,4	Palmeira (<i>Areca</i> spp.)		HC-Pro1 e HC-Pro2	(YANG <i>et al.</i> , 2018, 2019)
<i>Bevemovirus</i>	1	Monopartido	8,3	<i>Campanula takesimana</i>	Desconhecido	Carece da região P1, a HC-Pro carece dos motivos de transmissão por afídeos de potyvírus conservados R / KITC e PTK.	(SEO <i>et al.</i> , 2017)
<i>Brambyvirus</i>	1	Monopartido	10,8	Espécies de <i>Rubus</i>	Desconhecido	Domínio Alk1 codificado em uma região de codificação P1 muito grande. HC-Pro carece de motivos para amplificação do genoma e movimento sistêmico.	(SUSAIMUTHU <i>et al.</i> , 2008)
<i>Bymovirus</i>	6	Bipartido	RNA1: 7,2-7,6 RNA2: 2,2-3,6	Gramineas	<i>Polymyxa graminis</i>	Não têm regiões de codificação P1 e HC-Pro. RNA2 codifica um Proteína semelhante a HC-Pro.	CHEN <i>et al.</i> , 1999)
<i>Celavirus</i>	1	Monopartido	11,5	Ampla		5'-UTR inusualmente curta de 13 nt e 3'-UTR inusualmente longa de 586 nt.	(ROSE <i>et al.</i> , 2019)
<i>Ipomovirus</i>	7	Monopartido	9,2-10,8	Ampla	Mosca branca (<i>Bemisia tabaci</i>)	Alguns membros não possuem as regiões de codificação P1 e / ou HC-Pro e codifica uma proteína P1b em vez disso.	(JANSSEN <i>et al.</i> , 2005; LI <i>et al.</i> , 2008; MBANZIBWA <i>et al.</i> , 2011)
<i>Macluravirus</i>	10	Monopartido	8,2	Ampla	Afídeos	Sem região codificante de P1 e a região codificante de HC-Pro é mais curta do que em potyvírus.	(KONDO; FUJITA, 2012).
<i>Poacevirus</i>	3	Monopartido	9,7-10,2	Gramineae e Orchidaceae	Ácaro: triticum mosaic virus (TriMV)	5'-UTR excepcionalmente longo.	(WYLIE <i>et al.</i> , 2012; FELLERS <i>et al.</i> , 2009; TATINENI <i>et al.</i> , 2012)
<i>Potyvirus</i>	190	Monopartido	9,4-11,0	Ampla	Afídeos	ORF maior codifica para uma única poliproteína que é clivada em dez proteínas funcionais.	(VALLI; GARCÍA; LÓPEZ-MOYA, 2015; WYLIE <i>et al.</i> , 2017)

<i>Roymovirus</i>	2	Monopartido	9,5-9,9	<i>Passiflora edulis</i> , <i>Rosa sp.</i>	Possivelmente: ácaros eriófidos	Os motivos KITC e PTK estão ausentes no HC-Pro, assim como DAG no CP; o motivo C-2X-C no N-terminal de CP.	JOVER-GIL <i>et al.</i> , 2018; MOLLOV; LOCKHART; ZLESAK, 2013
<i>Rymovirus</i>	3	Monopartido	9,4-9,5	Gramineae	Ácaros eriófidos	ssRNA de 9,4-9,5 kb com organização genômica semelhante aos potyvírus.	(FRENCH; STENGER, 2005; WYLIE <i>et al.</i> , 2017; JO; CHOI; CHO, 2015)
<i>Tritimovirus</i>	6	Monopartido	9,2-9,6	Gramineae	Ácaros eriófidos	A proteína P1, em vez de HC-Pro, atua como um supressor de silenciamento de genes.	(STENGER <i>et al.</i> , 1998; VALLI; GARCÍA; LÓPEZ-MOYA, 2015; WYLIE <i>et al.</i> , 2017).

3.1. Organização genômica, expressão gênica e função das proteínas virais do gênero

Potyvirus

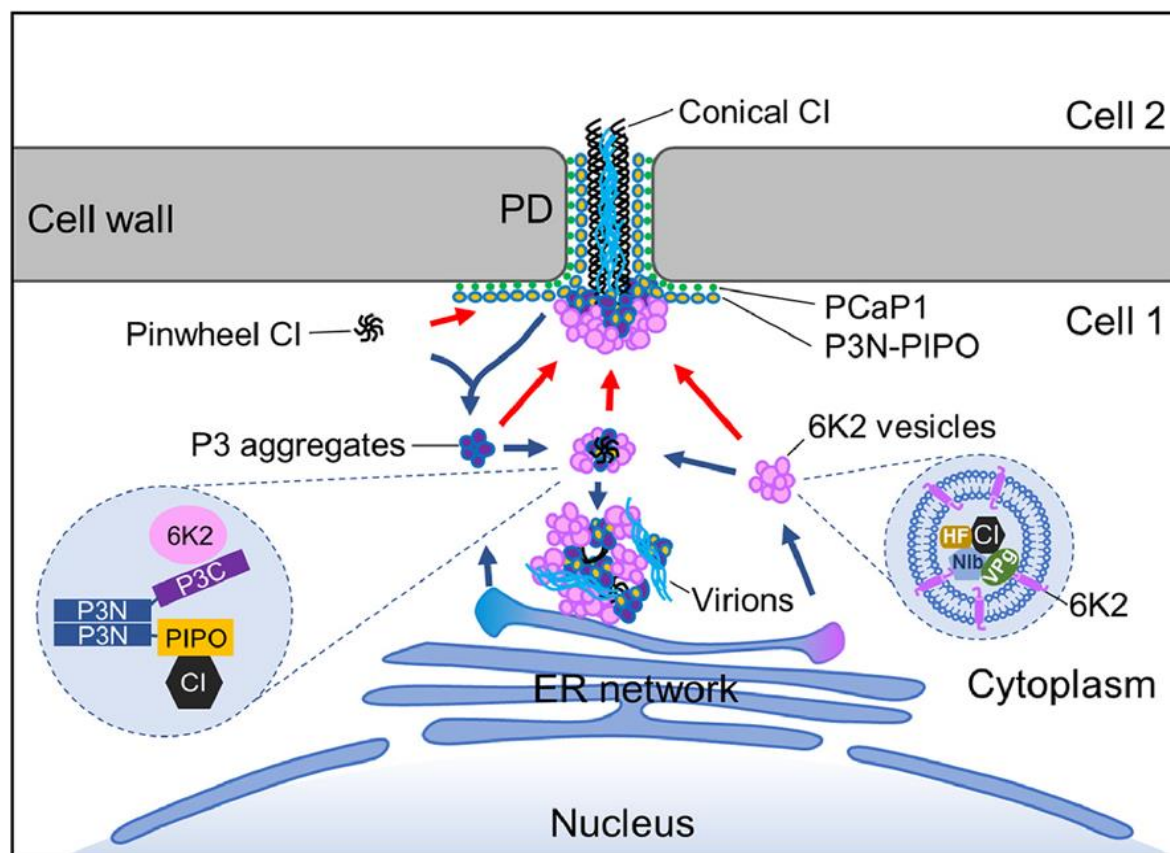
As partículas dos potyvírus são filamentosas ou flexuosas e não envelopadas (WYLIE *et al.*, 2017). A ORF maior codifica para uma única poliproteína que é clivada em dez proteínas funcionais (P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, e CP) por proteases auto-codificadas que reconhecem sítios semi-conservados. A proteína P1 (proteína 1 protease) é um tipo de serina-protease (S30). As proteínas HC-Pro e NIa-Pro são proteases do tipo cisteína C6 e C4, respectivamente (VALLI; GARCÍA; LÓPEZ-MOYA, 2015; WYLIE *et al.*, 2017). A P1 desempenha várias funções durante a infecção, incluindo a modulação da replicação viral e da tradução (MARTINEZ; DAROS, 2014; PASIN; SIMÓN-MATEO; GARCÍA, 2014). Martínez e Daròs (2015) observaram que no estágio inicial da infecção por tobacco etch virus (TEV) ocorre um acúmulo de P1 em células infectadas, especificamente dentro e fora do nucléolo. Além disso, a P1 pode se associar às subunidades ribossômicas 60S da planta tornando-as mais eficientes na tradução viral, uma vez que serve como um mediador no recrutamento das subunidades 60S no complexo iniciador da tradução viral. O domínio de protease é altamente conservado e localizado na região C-terminal da proteína, e a P1 dos potyvírus é a mais contrastante em tamanho (30-63 kDa) e também a mais divergente na sequência de aminoácidos (ADAMS; ANTONIW; FAUQUET, 2005; VALLI; GARCÍA; LÓPEZ-MOYA, 2015; YOSHIDA *et al.*, 2012). A região N-terminal hipervariável da P1 foi mapeada como um determinante de sintomas e está envolvida na patogenicidade e adaptação a respostas do hospedeiro (MALIOGKA *et al.*, 2012; PASIN; SIMÓN-MATEO; GARCÍA, 2014). Experimentalmente foi demonstrada a hipótese de que a proteína P1 estimula a atividade de supressão de silenciamento do RNA (ANANDALAKSHMI *et al.*, 1998; PRUSS *et al.*, 2004; RAJAMÄKI *et al.*, 2005; VALLI; LÓPEZ-MOYA; GARCÍA, 2007). Além disso, a P1 contribui para o aumento da infecção viral, mesmo em plantas com deficiência no mecanismo de silenciamento de RNA, sugerindo assim um papel independente da supressão do silenciamento de RNA (PASIN; SIMÓN-MATEO; GARCÍA, 2014). Uma outra pequena ORF (denominada PISPO) foi identificada na sequência de codificação da P1 de alguns potyvírus que infectam a batata-doce. A PISPO pode ser traduzida devido a uma mudança de quadro de leitura, semelhante à observada para a ORF PIPO (CLARK *et al.*, 2012; LI *et al.*, 2012). Não se sabe se a PISPO pode conferir alguma vantagem específica na infecção da batata-doce (REVERS; GARCÍA, 2015). No entanto, um estudo realizado por Mingot *et al.* (2016) sobre a expressão transitória dos produtos gênicos do sweet potato feathery mottle virus (SPFMV) na batata-doce revelou que a P1N-PISPO atua como um supressor de silenciamento de RNA e que mutações na

região que codifica o motivo WG/GW abolem a atividade de supressão de silenciamento. Com isso, Mingot et al. (2016) concluíram que sua função exigiria interação com os componentes “argonauta” da máquina de silenciamento. Pesquisas posteriores mostraram que a parte N-terminal de P1, P1-PISPO e HC-Pro de SPFMV têm atividade de supressão de silenciamento de RNA (RSS) (RODAMILANS; CASILLAS; GARCÍA, 2021; RODAMILANS *et al.*, 2018).

A proteína HC-Pro (*Helper Component Proteinase*) foi o primeiro supressor viral de silenciamento gênico identificado e está envolvida em movimento viral e na transmissão por pulgões (GOVIER; KASSANIS, 1974; KASSCHAU; CARRINGTON, 1998; SHIBOLETH *et al.*, 2007; VALLI *et al.*, 2014). Seis sítios essenciais foram localizados dentro de domínios putativos que envolvem a supressão do silenciamento do RNA e o mecanismo de defesa antiviral do hospedeiro, sendo este mecanismo crucial para a infecção e replicação viral (KASSCHAU; CARRINGTON, 2001). O motivo prolina-treonina-lisina (PTK) encontrado na proteína HC-Pro (ALA-POIKELA *et al.*, 2011; PENG *et al.*, 1998) é importante para a transmissão. Este domínio interage com o domínio altamente conservado DAG, localizado na região N-terminal da CP (PENG *et al.*, 1998; HUET *et al.*, 1994; BLANC *et al.*, 1997; LÓPEZ-MOYA; WANG; PIRONE, 1999; WYLIE *et al.*, 2017), facilitando assim a ligação da HC-Pro a CP do vírion (KIRAN *et al.*, 2020). O motivo KITC localizado na região N-terminal de HC-Pro está envolvido na retenção do vírus no estilete de pulgões (ATREYA *et al.*, 1992; ATREYA; PIRONE, 1993; PENG *et al.*, 1998) em transmissões do tipo não-persistente (GIBBS *et al.*, 2008). O motivo FRNK, altamente conservado, localizado na região central de HC-Pro está envolvido na supressão de silenciamento de RNA e está associado ao desenvolvimento de sintomas e aumento da carga viral (SHIBOLETH *et al.*, 2007; TORRANCE *et al.*, 2006). A CP, além de ser uma proteína de revestimento que estruturalmente forma parte dos vírions, está envolvida na transmissão por pulgões e no movimento viral (CUI; WANG, 2016; WYLIE *et al.*, 2017). A proteína potyviral 3 (P3) participa da replicação, movimento, patogenicidade do vírus (JENNER *et al.*, 2003; JOHANSEN *et al.*, 2001; SUEHIRO *et al.*, 2004) e está envolvida na adaptação do vírus ao hospedeiro (NIGAM *et al.*, 2019; WEN; MAROOF; HAJIMORAD, 2011). O cistron P3 codifica duas proteínas virais, P3 e P3N-PIPO, que possuem funções diferentes (SHI *et al.*, 2020). A proteína P3 possui duas regiões hidrofóbicas, uma localizada na região N-terminal e outra na extremidade C-terminal (KEKARAINEN; SAVILAHTI; VALKONEN, 2002). A extremidade C-terminal é responsável por direcionar a P3 para a membrana do retículo endoplasmático (RE) e formar inclusões associadas ao aparelho de Golgi, que transitam ao longo dos filamentos de actina e co-localizam com as vesículas de replicação denominadas de complexos de replicação viral associada à membrana (VRC) (KEKARAINEN; SAVILAHTI;

VALKONEN, 2002; CUI *et al.*, 2017). A proteína P3N-PIPO interage com a P3 através da região N-terminal compartilhada, sendo essa interação essencial para o movimento viral célula-a-célula (CHUNG *et al.*, 2008; OLSPERT *et al.*, 2015; RODAMILANS *et al.*, 2015; WEI; ZHANG *et al.*, 2010; WHITE, 2015), e possivelmente para o recrutamento de vesículas induzidas por 6K2 para o plasmodesmo (PD) (CHAI *et al.*, 2020). Chai *et al.* (2020) propõem um modelo para o movimento potyviral de célula-a-célula (Figura 3), indicando que agregados de P3 derivados de ER são recrutados para formar agregados com 6K2 através da interação entre o domínio P3C e 6K2. A P3N-PIPO localizada na membrana plasmática e no PD são recrutados pela P3 para os agregados de 6K2. Por outro lado, a P3N-PIPO em associação com a PCaP1 (proteína 1 de ligação de Ca^{2+} associada à membrana plasmática (PM)) recruta a proteína CI para o PD, enquanto a P3 co-localizada com a P3N-PIPO recruta a CI para o agregado de 6K2. Além disso, moléculas protéicas de CI unem-se ao complexo CI/P3N-PIPO através de processos de auto-interação para formarem estruturas cônicas no PD ou estruturas de cataventos no citoplasma. Por fim, os agregados citoplasmáticos com 6K2 agrupam-se com as proteínas P3, P3N-PIPO e CI na área perinuclear para formar um viroplasma irregular típico induzido por vírus ou são enviados para o PD e ancorados em inclusões cônicas contendo a proteína CI. No PD, os RNAs genômicos virais recém-sintetizados são totalmente ou parcialmente encapsulados por CP em vírions ou complexos ribonucleoproteicos (RNP) e, em seguida, transportados com auxílio da CI para células vizinhas (CHAI *et al.*, 2020).

Figura 3 - Modelo esquemático do movimento célula-a-célula do potyvírus turnip mosaic virus (TuMV).



CI, P3N-PIPO, P3, 6K2, vírions, PCaP1, VPg, Nib e fator de hospedeiro (HF) são representados em cataventos pretos ou hexágonos, esferas amarelas com margens azuis, esferas roxas com margens azuis, esferas ou cilindros rosa, linhas ciano, pontos verdes, um oval verde escuro, e um pentágono azul, respectivamente. As setas vermelha e azul escuro indicam a possível rota para a montagem de complexos CI / P3N-PIPO / P3 / 6K2 induzidos por vírus citoplasmáticos e localizados no plasmodesmo (PD), respectivamente. As duas áreas pontilhadas mostram a rede de interação entre 6K2, P3, P3N-PIPO e CI e uma vesícula 6K2 replicativa típica, respectivamente. O núcleo, citoplasma, parede celular, rede de retículo endoplasmático (ER) e PD também são indicados. Nem todos os elementos são desenhados em escala. Retirado de Chai et al. (2020).

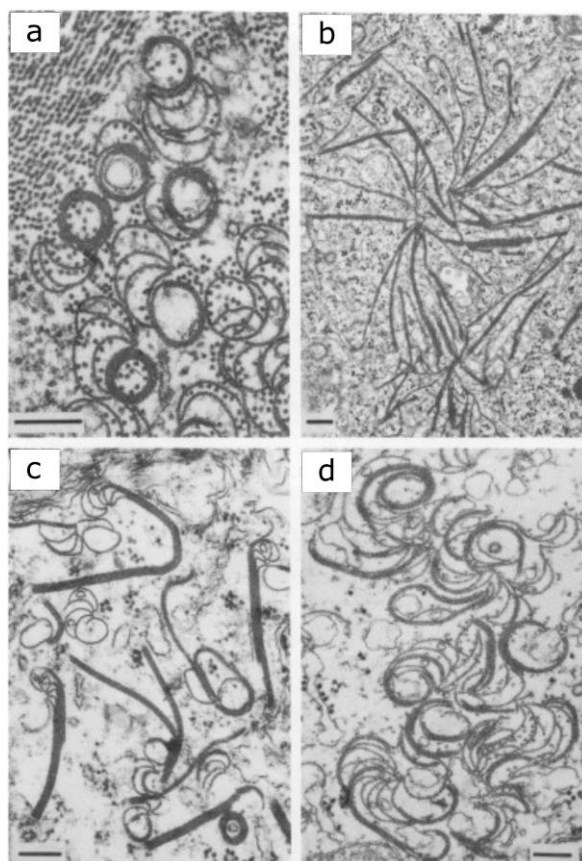
A curta ORF PIPO (*Pretty Interesting Potyviridae ORF*) curta é expressa através do deslizamento da RNA polimerase dependente de RNA (RdRp) no sítio conservado G₁₋₂A₆₋₇ (G) GAA AAA A (A) (os espaços separam os códons de poliproteína) localizado na extremidade 5' da ORF PIPO (CHUNG *et al.*, 2008; OLSPERT *et al.*, 2015). Este sítio GA₆ e outro An (n >= 6) podem variar entre as sequências virais de membros da família *Potyviridae*. Por exemplo, em genomas das espécies do gênero *Bymovirus*, o sítio é GGA AAA AA, e em sequências genômicas de espécies do gênero *Tritimovirus* a sequência pode ser G GAA AAA ou GGA AAA AAA (CHUNG *et al.*, 2008; RODAMILANS *et al.*, 2015). Segundo Rodamilans et al. (2015) a presença de adições de resíduo A no motivo PIPO GA₆ ocorre em 1,6% das leituras transcritas.

O peptídeo 1 (6K1) e o peptídeo 2 (6K2) de seis quilodaltons são as duas menores proteínas potyvíricas. Os peptídeos 6K1 e 6K2 flanqueados pelo domínio CI contém extensões

hidrofóbicas e estão envolvidas na replicação e na ligação à membrana (GIACHETTI; SEMLER, 1991; JOHNSON; SARNOV, 1991). A 6K1 é necessária para a replicação do vírus (CUI; WANG, 2016), e sua função ainda precisa ser melhor caracterizada. No entanto, experimentos realizados por Cui e Wang (2016) revelaram que deleções pequenas de qualquer região gênica de 6K1 (de 5 a 13 aminoácidos) ou da sequência de 6K1 sem considerar os sítios de clivagem, ou da sequência completa de 6K1 do clone infeccioso de plum pox virus (PPV), abolem a replicação viral. A proteína 6K2 induz à formação de vesículas virais nos locais de saída do retículo endoplasmático (ER), e são direcionadas aos cloroplastos para a replicação viral robusta (WEI *et al.*, 2010; WEI; WANG, 2008; WYLIE *et al.*, 2017). O domínio hidrofóbico localizado na parte central de 6K2 está relacionado à ligação à membrana necessária ao processo de infecção viral (SCHAAD; JENSEN; CARRINGTON, 1997).

A proteína de inclusão citoplasmática (CI) é uma helicase envolvida no movimento e na replicação viral (FERNÁNDEZ *et al.*, 1997; GABRENAITE-VERKHOVSKAYA *et al.*, 2008; WEI; ZHANG *et al.*, 2010), e possui atividade de ATPase (EAGLES *et al.*, 1994). Como já mencionado, inclusões cilíndricas na forma de cata-ventos típicos são localizadas no citoplasma de células infectadas com potyvírus (EDWARSON; CHRISTINE; KO, 1984) (Figura 4). Como citado anteriormente, a proteína CI atua em colaboração com a P3N-PIPO para ajudar no movimento viral (GÓMEZ DE CEDRÓN *et al.*, 2006; WEI; ZHANG *et al.*, 2010). Além disso, a CI atua como um fator de virulência para diferentes genes de resistência (SOREL; GARCIA; GERMAN-RETANA, 2014).

Figura 4 - Modificações morfológicas de inclusões citoplasmáticas.



De acordo com a subdivisão de Edwarson (1966) e Edwarson, Christie e Ko (1984). **a.** espirais induzidos em *Nicotiana clevelandii* por um isolado de berinjela não identificado. **b.** agregados laminados induzidos pelo vírus Y asiático em plantas de *Chenopodium quinoa*. **c.** espirais induzidos por TuMV e agregados laminados em *N. clevelandii*. **d.** espirais agregados curtos e curvos em *N. tabacum* infectados por PVY. Barras= 200 nm

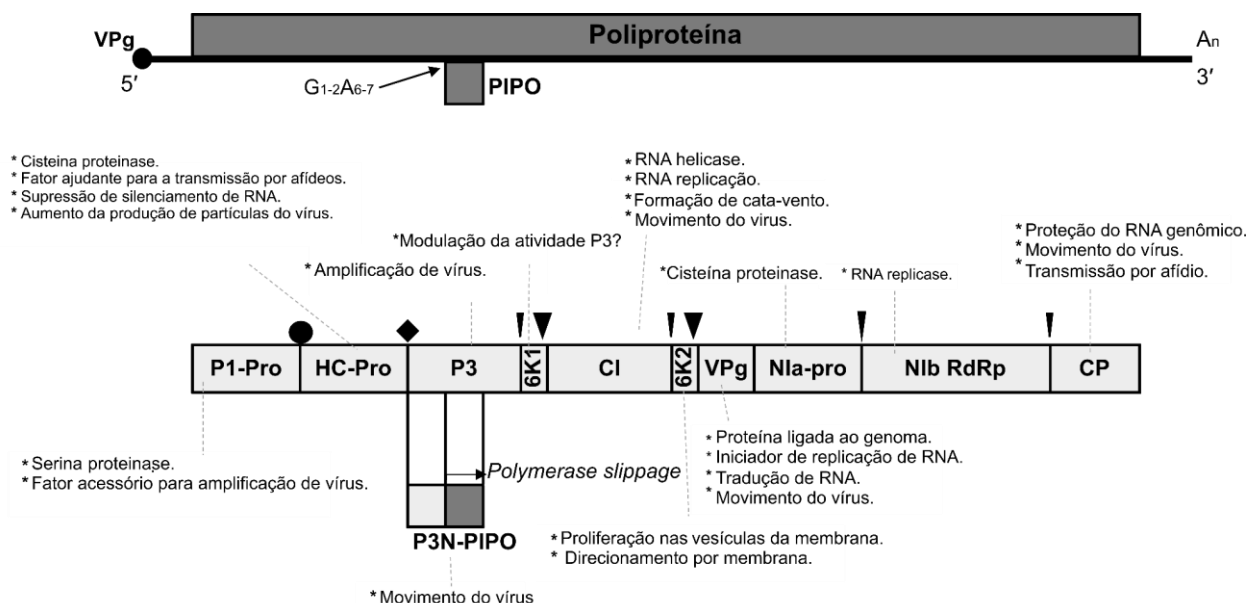
A protease de inclusão nuclear A (NIa-Pro) está envolvida no processamento de poliproteínas, enquanto a inclusão nuclear B (NIb) é a RdRp responsável pela replicação do genoma viral (SHEN *et al.*, 2020). A NIa-Pro é a principal protease viral responsável pela clivagem da maioria dos sítios da poliproteína (SUN *et al.*, 2010). A NIb, além de seu papel principal como uma replicase, recruta várias proteínas do hospedeiro para os complexos de replicação viral (VRCs), e interage com outras proteínas para suprimir a resposta de imunidade (SHEN *et al.*, 2020).

A proteína ligada ao genoma (VPg) é uma proteína relativamente pequena (~ 21 kDa) que está covalentemente ligada ao RNA viral através de uma ligação fosfodiéster entre um resíduo conservado de serina ou tirosina com o terminal 5' da uridina do nucleotídeo do RNA viral (extremidade 5' viral) (MURPHY *et al.*, 1996; DREHERA; MILLERB, 2006). A proteína é intrinsecamente desordenada, e por ser desordenada ou com flexibilidade estrutural pode formar complexos de interação com diferentes proteínas para funções diversas, interagindo, portanto, com múltiplos parceiros (CHENG; WANG, 2017). A VPg-NIa é constituída pela

metade da proteína NIa (N-proximal), e a outra metade da região C-proximal correspondente ao domínio da protease (NIa-Pro), os quais são separados por um sítio por meio de uma incisão proteolítica subótima (KEKARAINEN; SAVILAHTI; VALKONEN, 2002). A VPg está envolvida na tradução, movimento e replicação (ESKELIN *et al.*, 2011; REVERS; GARCÍA, 2015). Além disso, a maioria dos potyvírus codifica dois supressores de silenciamento de RNA viral (RSS), um deles a HC-Pro, e o outro a proteína VPg (CHENG; WANG, 2017). No entanto, há casos em que outros genes podem desempenhar a função de RSS, como é o caso da P1N-PISPO do potyvírus sweet potato feathery mottle potyvirus (MINGOT *et al.*, 2016). Como já mencionado, nos tritimovírus, a proteína P1, em vez de HC-Pro, também atua como um supressor de silenciamento gênico (STENGER *et al.*, 1998; VALLI; GARCÍA; LÓPEZ-MOYA, 2015; WYLIE *et al.*, 2017). Um grande número de genomas virais de ssRNA que infectam plantas, incluindo os potyvírus, tem VPg em vez da estrutura 5' quepe (*cap*) (DREHER; MILLER, 2006). Os RNAs virais ligados a VPg geralmente não possuem 5'-UTRs longos altamente estruturados com AUGs integrados, que forneçam uma tradução independente de 5' *cap*. Na ausência do 5' *cap*, diversos elementos de estrutura secundária de RNA estão presentes nas sequências das regiões 5' e 3' não traduzidas que recrutam o complexo ribossomal e iniciam a tradução a partir do 5' AUG proximal (BELSHAM; SONENBERG, 2000; ZHANG; ROBERTS; RAKOTONDRAFARA, 2015). Portanto, a poliproteína viral dos potyvírus é traduzida por meio de um mecanismo independente de *cap* (BELSHAM; SONENBERG, 2000; NIEPEL; GALLIE, 1999). A região 5'-UTR carece de uma estrutura secundária estável previsível, no entanto, apresentam elementos regulatórios independentes que promovem o início da tradução (KEKARAINEN *et al.*, 1999; SIMÓN-BUELA; GUO; GARCÍA, 1997).

As características e funções relevantes das proteínas de potyvírus estão resumidas na Figura 5 e na Tabela 4.

Figura 5 - Mapa genômico de um membro típico do gênero *Potyvirus*. Características relevantes e funções propostas.



O genoma ssRNA é representado por uma linha e a poliproteína ORF por um quadro de leitura aberta com os produtos proteolíticos maduros nomeados. VPg (proteína viral ligada ao genoma), ligada covalentemente ao nucleotídeo 5' terminal, é representada por um círculo; a cauda poli (A) na região 3' terminal é representado por A_n; P1-Pro (proteína 1 protease), uma proteína com atividade proteolítica de serina responsável pela clivagem tipicamente em sítios de Tyr/Phe-Ser (●); HC-Pro (componente auxiliar protease), é uma proteína envolvida na transmissão por pulgões e atividade proteolítica de cisteína responsável pela clivagem nos sítios Gly-Gly (◆); P3 (proteína 3); PIPO (*pretty interesting Potyviridae ORF*); 6K (péptido de seis quilodaltos); CI (proteína de inclusão citoplasmática); NIa-Pro (protease de inclusão A nuclear), com atividade proteolítica semelhante à cisteína responsável pela clivagem em sítios Gln/Glu- (Ser/Gly/Ala) (▼); NIb (proteína de inclusão nuclear B, RNA polimerase dependente de RNA); CP (proteína de revestimento). Os locais de clivagem de P1-Pro, (●), HC-Pro (◆) e NIa-Pro (▼) são indicados.

Adaptado: ICTV (2021) (https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/potyviridae); ViralZone (<https://viralzone.expasy.org/48>), OLSPERT *et al.* (2015); REVERS; GARCÍA, (2015) e RODAMILANS *et al.* (2015).

Tabela 4 - Função das proteínas dos vírus do gênero *Potyvirus*.

Proteína	Função	Fonte bibliográfica
P1	Tradução, movimento viral pela atividade de ligação ao RNA, replicação do genoma viral e atividade protease.	(MARTINEZ; DAROS, 2014; REVERS; GARCÍA., 2015; VALLI; GARCÍA; LÓPEZ-MOYA, 2015; WYLIE <i>et al.</i> , 2017)
HC-Pro	Transmissão por afídeos, desenvolvimento de sintomatologia; movimento célula-a-célula, e supressão de silenciamento gênico.	(GOVIER; KASSANIS, 1974; KASSCHAU; CARRINGTON, 1998; SHIBOLETH <i>et al.</i> , 2007; TORRANCE <i>et al.</i> , 2006; VALLI <i>et al.</i> , 2014)
P3	Determinante da patogenicidade, replicação e movimento pela ligação com o complexo viral.	(NIGAM <i>et al.</i> , 2019; WEN; MAROOF; HAJIMORAD, 2011)

P3N-PIPO	Movimento célula-a-célula.	(CHUNG <i>et al.</i> , 2008; OLSPERT <i>et al.</i> , 2015; REVERS; GARCÍA, J.A., 2015; WEI; ZHANG <i>et al.</i> , 2010; WHITE, 2015)
6K1	Formação de vesículas de replicação.	(CUI; WANG, A., 2016; WYLIE <i>et al.</i> , 2017)
CI	Atividade helicase, envolvida no movimento viral e replicação.	(FERNÁNDEZ <i>et al.</i> , 1997; GABRENAITE-VERKHOVSKAYA <i>et al.</i> , 2008; WEI; ZHANG <i>et al.</i> , 2010)
6K2	Formação de vesículas de replicação.	(CUI; WANG, 2016; WYLIE <i>et al.</i> , 2017)
VPg	Tradução, movimento e replicação viral.	(ESKELIN <i>et al.</i> , 2011; REVERS; GARCÍA, 2015)
NIa-Pro	Processamento da poliproteína.	(SUN <i>et al.</i> , 2010).
NIb	Atividade de translocação nuclear; RNA polimerase dependente de RNA.	(CUI; WANG, 2016; SHEN <i>et al.</i> , 2020; WYLIE <i>et al.</i> , 2017)
CP	Transmissão por afídeos, movimento célula-a-célula e de longa distância, formação do virion.	(CUI; WANG, 2016; VENTER <i>et al.</i> , 2001; WYLIE <i>et al.</i> , 2017; YAN <i>et al.</i> , 2021).

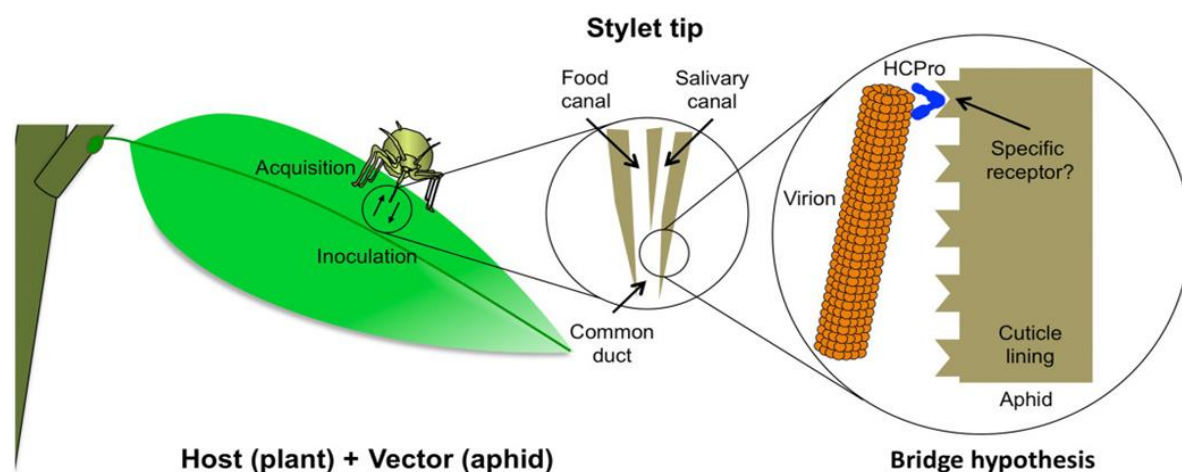
De acordo com o ICTV, o critério de demarcação para espécies dentro do gênero *Potyvirus* estipula 76% de identidade à nível de nucleotídeos e mais de 82% de identidade à nível de aminoácidos (aa) para a ORF que codifica a poliproteína, enquanto 76-77 % identidade (nt) para sequências de CP, e pelo menos 58% de identidade nucleotídica para P1 e entre 74-78% de identidade nucleotídica para outros genes individuais (ADAMS; ANTONIW; FAUQUET, 2005).

3.2. Transmissão

Membros do gênero *Potyvirus* são normalmente transmitidos por pulgões de forma não-persistente e não-circulativa, e os motivos conservados KITC, PTK (localizado na região de codificação da HC-Pro) e DAG (localizado na região de codificação da CP) são necessários para a transmissão por esses insetos vetores (KIRAN R. *et al.*, 2020; NG; PERRY, 2004; PIRONE; PERRY, 2002; WYLIE *et al.*, 2017). Além disso, os vírus deste gênero podem ser transmitidos por inoculação mecânica, por enxertia (MCKERN *et al.*, 2008), por sementes (JOSSEY; HOBBS; DOMIER, 2013; REVERS; GARCÍA, 2015) e pólen (CARD; PEARSON; CLOVER, 2007).

Na transmissão do tipo não-persistente, o inseto vetor adquire o vírus em segundos ou minutos e o pulgão permanece virulífero somente por poucas picadas de prova (KIRAN *et al.*, 2020). A interação dos vírions ao estilete do pulgão é realizada por meio de duas estratégias (FROISSART; MICHALAKIS; BLANC 2002): a estratégia do capsídeo, onde a CP interage diretamente com o receptor ou com o ponto de ligação no estilete do pulgão, e a estratégia auxiliar, onde a HC-Pro facilita a ligação entre a CP e o receptor através de uma “ponte molecular” (*Bridge Hypothesis*) reversível (BLANC *et al.*, 1997; DOMBROVSKY *et al.*, 2007; KIRAN *et al.*, 2020; VALLI *et al.*, 2018) (Figura 6). Essa interação íntima entre a HC-Pro e a CP permite a interação do vírion com o estilete do pulgão (DIETZGEN; MANN; JOHNSON, 2016; VALLI *et al.*, 2018). O motivo DAG interage diretamente com o motivo PTK (HUET *et al.*, 1994; PENG *et al.*, 1998), e o motivo KITC e/ou os motivos funcionalmente equivalentes localizados na extremidade N-terminal da HC-Pro são essenciais para a retenção do vírus no estilete dos pulgões (BLANC *et al.*, 1998; HUET *et al.*, 1994). Wylie *et al.* (2017) demonstraram que os potyvírus sem esses três motivos não são transmitidos por afídeos. Em um estudo realizado por Blanc *et al.* (1997) com algumas variantes do vírus da mancha das nervuras do tabaco (tobacco vein mottling virus, TVMV), observou-se que sete aminoácidos na CP que incluem o motivo DAG (DTVVDAGK) estão envolvidos na transmissibilidade do pulgão e sua capacidade de ligação a HC-Pro. A caracterização dos receptores da HC-Pro em pulgões ainda é pouco elucidada (FERNÁNDEZ-CALVINO *et al.*, 2010), porém, há estudos (DOMBROVSKY *et al.*, 2007; FERNÁNDEZ-CALVINO *et al.*, 2010) onde são mencionados a existência de proteínas cuticulares no estilete do inseto como receptores da HC-Pro. Por exemplo, Fernández-Calvino *et al.* (2010) identificaram nove proteínas extraídas da cabeça de afídeos que interagem com a HC-Pro de TEV, entre estas, a proteína ribossomal S2 (PRS2).

Figura 6 - ‘Hipótese ponte’ para transmissão de potyvírus por pulgões.



Esquerda: um pulgão está se alimentando de uma planta infectada. **Centro:** seção longitudinal do estilete mandibular (os flancos externos da maxila foram omitidos para simplificar a figura), incluindo os dois canais paralelos (o canal alimentar que se conecta ao sistema digestivo e o canal salivar que permite as secreções durante a alimentação) unindo-se no duto comum. **À direita:** um complexo de proteinase componente auxiliar (HC-Pro) (representado em uma forma dimérica) é ligado a uma extremidade da partícula viral e permite uma interação reversível com receptores potenciais localizados sobre o revestimento da cutícula (lado interno da ponta do estilete). Deve-se notar que esta figura é uma representação preditiva do processo de transmissão viral com base em dados experimentais disponíveis muito limitados sobre interações e o papel consequente do HC-Pro durante este processo (ver texto para detalhes). Portanto, não se pode excluir, por exemplo, que a interação HC-Pro-CP pode ocorrer ao longo de toda a partícula viral e as formas não diméricas de HC-Pro podem desempenhar um papel na transmissão viral (retirado de VALLI *et al.*, 2018).

3.3. Transmissão por afídeos

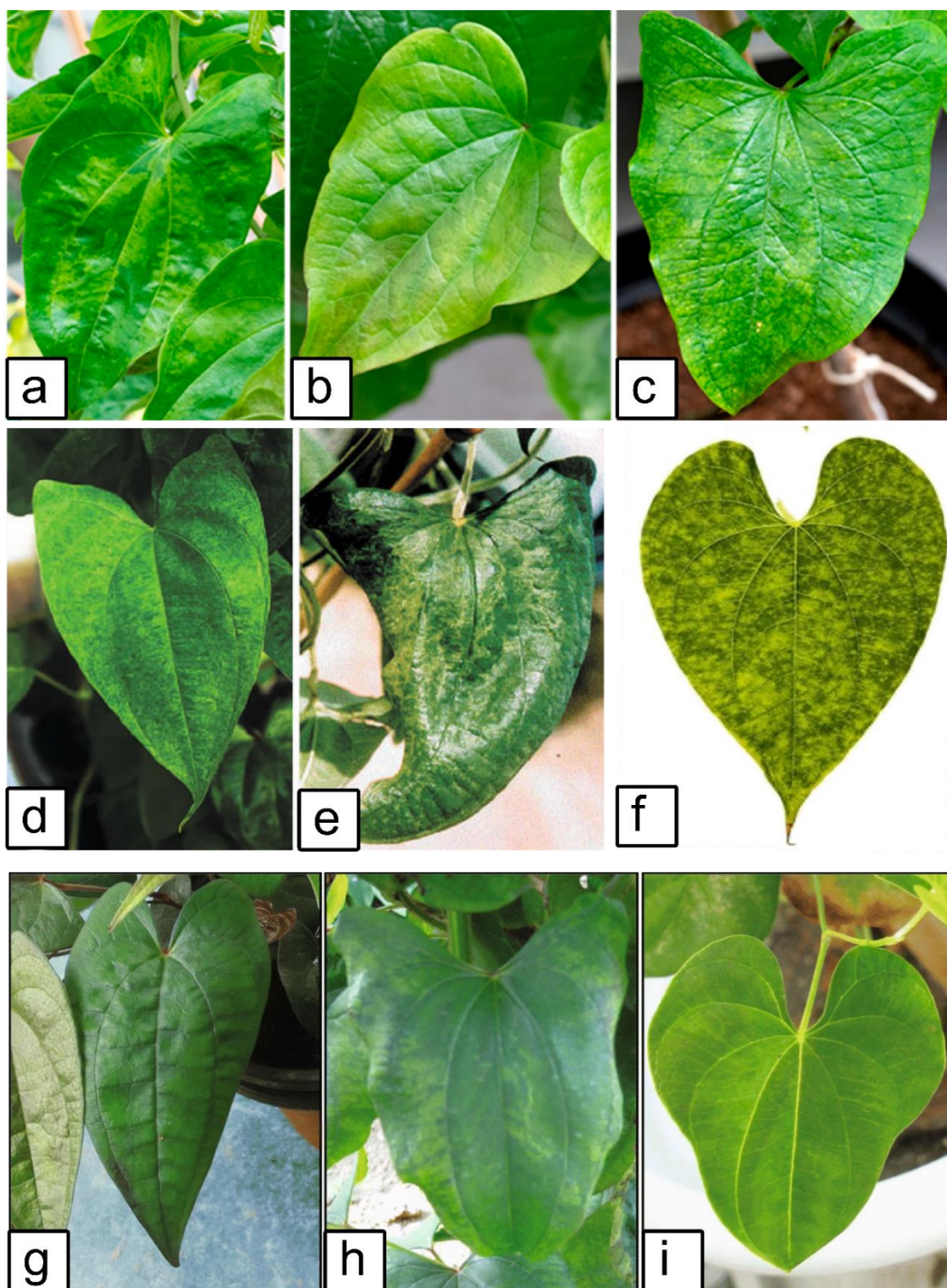
Os potyvírus dependem de pulgões para o movimento de planta-a-planta e transmissão rápida. No entanto, a relação com o vetor é transitória e de curta duração (BRAGARD *et al.*, 2013; DIETZGEN; MANN; JOHNSON, 2016; HOGENHOUT *et al.*, 2008; NG; PERRY, 2004), onde o vírus pode ser adquirido entre 3-5 segundos durante a penetração do estilete na célula epidérmica da planta (MORENO *et al.*, 2012; POWELL, 2005; POWELL; PIRONE; HARDIE, 1995). Kiran *et al.* (2020) realizaram uma análise filogenética usando sequências de nucleotídeos da CP de 176 potyvírus. Os autores revelaram que a maioria dos 176 potyvírus foram transmitidos por menos de três espécies de afídeos ou não tinham vetores conhecidos. De acordo com a análise metagenômica de 137 espécies de afídeos, observou-se que 45% eram polípagos, 38% oligófagos e apenas 17% monófagos (KIRAN R. *et al.*, 2020) e o vetor mais eficaz dentro das dez espécies principais foi o *Myzus persicae*, que é altamente polífago e cosmopolita. As outras nove espécies corresponderam a *Aphis gossypii*, *A. craccivora*, *A. fabae*, *A. spiraecola*, *A. pisum*, *Macrosiphum euphorbiae*, *Rhopalosiphum maidis*, *R. padi* e *Lipaphis erysimi*. Portanto, é de se esperar que os pulgões transmitam múltiplos potyvírus com diferentes eficiências e diversas estratégias de disseminação viral (KIRAN *et al.*, 2020). Não se sabe com

exatidão se os vírus transmitidos de forma não-persistentes estão intimamente associados a insetos vetores (LEFEUVRE *et al.*, 2019), e, portanto, a eficácia na transmissão parece depender do comportamento colonizador do vetor (FERERES; MORENO, 2009; HALBERT; IRWIN; GOODMAN, 1981; POWELL; TOSH; HARDIE, 2006; STAFFORD; WALKER; ULLMAN, 2012), visto que, aqueles pulgões que mal realizam a picada de prova e não colonizam a planta parecem ser os melhores transmissores de potyvírus (FERNÁNDEZ-CALVINO *et al.*, 2006; HALBERT; IRWIN; GOODMAN, 1981; KALLESHWARASWAMY; KRISHNA KUMAR, 2008; PEREZ *et al.*, 1995). Além disso, os potyvírus não são especializados em um único tipo de tecido vegetal (RESTREPO; FREED; CARRINGTON, 1990), e por isso, podem ser facilmente adquiridos e inoculados na fase inicial de punção intracelular do estilete (MARTIN *et al.*, 1997; MORENO *et al.*, 2012). Deve-se considerar também a importância dos compostos orgânicos voláteis (COV) da planta, que, quando alterados, podem levar à imigração ou emigração do vetor para a planta infectada com o vírus (KIRAN *et al.*, 2020). Eigenbrode *et al.* (2002) compararam a permanência de *M. persicae* entre plantas infectadas com PVY e não infectadas e observaram que houve um aumento na retenção de pulgões e significativamente uma menor emigração de plantas infectadas. Além disso, Eckel e Lampert (1996) relataram uma maior atração de pulgões por plantas de tabaco infectadas com TEV. Em contraste, Boquel, Giordanengo e Ameline (2011) relataram uma redução na preferência de *M. euphorbiae* por batata infectada com PVY. Os potyvírus também podem tornar as plantas palatáveis ou nutritivas para o inseto vetor, modificando a aptidão alimentar pela planta (CASTLE; BERGER, 1993; WESTWOOD *et al.*, 2014). Alguns exemplos são observados na literatura. O vetor *A. gossypi* demonstra ter maior longevidade em plantas de abóbora infectadas com o vírus do mosaico amarelo da abobrinha (ZYMV) (BLUA; PERRING; MADORE, 1994). O aumento da população de *L. erysimi* em plantas infectadas com TuMV foi demonstrado por Adachi *et al.* (2018), e um maior crescimento, reprodução e sobrevivência de *M. persicae* em plantas de batata infectadas com PVY por Castle e Berger (1993). A interação hospedeiro-vetor mediada por potyvírus é governada em parte por alterações na expressão gênica, resultando na alteração da biossíntese de sinais olfativos e gustativos (BAK *et al.*, 2019; CASTEEL *et al.*, 2014; WESTWOOD *et al.*, 2014). A indução da produção de salicilatos em tomateiros infectados com PVY, que leva ao aumento da fertilidade de *M. euphorbiae*, foi observada por Kersch-Becker e Thaller (2014). Por último, Bak *et al.* (2019) demonstraram que a sinalização do etileno em plantas infectadas com PVY tornam estas mais atrativas para os pulgões (BAK *et al.*, 2019).

3.4. Sintomas e sinais

As interações vírus-hospedeiro compatíveis ou parcialmente compatíveis afetam a fisiologia da planta e, em alguns casos, induzem uma variabilidade de sintomas (REVERS; GARCÍA, 2015). Os potyvírus comumente causam uma ampla variação de sintomas (Figura 7) que provocam doenças e perdas econômicas significativas. Os sintomas macroscópicos podem ser resumidos principalmente em mosaicos, manchas, retardo de crescimento, arranhões ou estrias, clorose, clareamento de nervuras, necrose, entre outros, e ocasionalmente são observadas alterações nas flores, frutos ou sementes (VALLI; GARCÍA; LÓPEZ-MOYA, 2015). É muito comum o efeito de sinergismo com os potyvírus, onde expressões de sintomas mais graves podem ocorrer (GONZÁLEZ-JARA *et al.*, 2004; VALLI; GARCÍA; LÓPEZ-MOYA, 2015). Isso pode estar associado à atividade cruzada do supressor de silenciamento viral que afeta o processo de infecção do vírus não relacionado (PRUSS *et al.*, 1997) e/ou a capacidade de P3N-PIPO facilitar o movimento viral através dos plasmodesmos (HISA *et al.*, 2014).

Figura 7 - Alguns exemplos de sintomas em espécies de dioscóreas causados por potyvírus

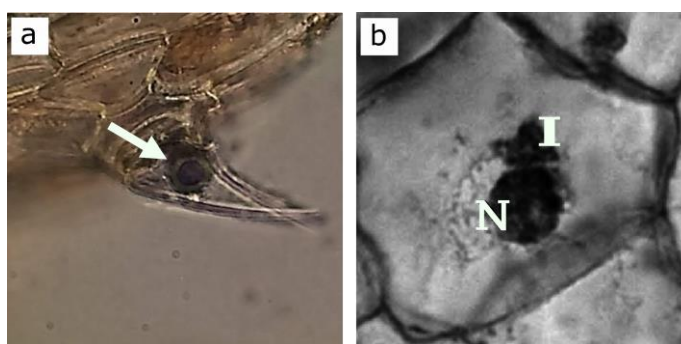


Plantas de *Dioscorea cayennensis-rotundata* que testaram positivo para YMV com sintomas de mosaico (a), descoloração clorótica das folhas (b) e manchas (c). Folhas de *D. alata* infectadas com YMMV (d) e YMV (e). Folhas de *D. bulbifera* apresentando sintomas de mosaico infectadas com dioscorea mosaic virus (*Potyvirus*) e air potato virus 1 (*Ampelovirus*) (f). Inhames infectados apenas por YMMV, *D. japonica* (g); *D. alata* (h) e *D. oposita* (i).

Fonte: a-c: Silva et al. (2019); d-e: Adams et al. (2003); f: Dey et al. (2019); g-i: Zou et al. (2020)

Em células infectadas pode ocorrer a presença de corpos de inclusão citoplasmática (Figura 8), bem como a presença de partículas virais no citoplasma e anormalidades em organelas (VALLI; GARCÍA; LÓPEZ-MOYA, 2015). Os potyvírus induzem inclusões cilíndricas citoplasmáticas que consistem na forma de cataventos, pergaminhos e agregados laminados curtos e curvos (Figura 4) (EDWARDSON, 1966, 1992; EDWARSON; PURCIFULL; CHRISTINE, 1968, 1984). As células infectadas também podem apresentar inclusões do tipo nucleares cristalinos (CARRINGTON; DOUGHERTY, 1987). Por exemplo, o RNA de TEV codifica proteínas que se agregam dentro dos núcleos das células infectadas que formam inclusões nucleares em forma de placa (DOUGHERTY; HIEBERT, 1980)

Figura 8 - Corpos de inclusão cristalinos em tecidos vegetais infectados com potyvírus.



Tricomas de *Passiflora edulis* infectados com o vírus do mosaico do caupi transmitido por afídeo (CABMV) **(a)** (RISCO e BLAWID, 2019). Células do tecido foliar de *Brugmansia* (**I** = inclusão cilíndrica; **N** = núcleo) coradas com laranja-verde **(b)** (CHELLEMI *et al.*, 2011).

Ambos, tanto as sequências 5' e 3'-UTR quanto os produtos gênicos dos potyvírus (P1, HC-Pro, P3, 6K1, CI, 6K2 e VPg) estão envolvidos na expressão de sintomas (VALLI; GARCÍA; LÓPEZ-MOYA, 2015). Mutações na HC-Pro podem causar alterações drásticas na expressão de sintomas (HAIKONEN *et al.*, 2013; SHIBOLETH *et al.*, 2007). A supressão do silenciamento pode afetar os microRNAs (miRNAs) que estão envolvidos no desenvolvimento e outras funções essenciais da planta. Os miRNAs podem alterar a expressão de alguns sintomas virais (KASSCHAU *et al.*, 2003), bloquear a síntese de proteínas endógenas (ARANDA; MAULE, 1998; BALLUT *et al.*, 2005). Muitos supressores de silenciamento virais funcionam como fatores de patogenicidade que causam sintomas ou anormalidades no desenvolvimento da planta (CHAPMAN *et al.*, 2004). Pequenos RNAs de interferência (siRNAs) são envolvidos no silenciamento gênico, e os micro-RNAs endógenos estão envolvidos na regulação da expressão gênica em plantas (CHAPMAN *et al.*, 2004). No entanto, proteínas virais podem suprimir o mecanismo de silenciamento do RNA como um mecanismo de contra-resposta que permite a infecção sistêmica viral na planta (CHAPMAN *et al.*, 2004). Portanto, proteínas supressoras de vírus inibem a atividade de siRNAs/miRNAs e desencadeiam respostas que podem ser expressas

em alterações no desenvolvimento da planta (CHAPMAN *et al.*, 2004). Por exemplo, a proteína HC-Pro afeta a via de siRNA e miRNA como meio de controle da expressão gênica no hospedeiro (MALLORY *et al.*, 2002). A HC-Pro bloqueia o mecanismo de silenciamento de RNA em plantas, seja por inibição do processamento ou desestabilização de siRNA, fazendo com que as quantidades de siRNA e miRNA em plantas possam ser reguladas diferencialmente pela proteína viral (MALLORY *et al.*, 2002). Kasschau *et al.* (2003) observaram que altos níveis de supressores virais (P1 e HC-Pro de TuMV) interferiram na atividade de miRNA envolvidos na degradação de vários mRNA que codificam fatores de transcrição em plantas (KASSCHAU *et al.*, 2003).

Danos específicos podem ser causados pelos potyvírus: o PVY pode causar sintomas sistêmicos de necrose nas nervuras de tabaco (JONES, 1990; JONES; CENTER, 2018), mosaicos, deformação das folhas e/ou atraso de crescimento em batateiras (FUENTES *et al.*, 2019). O PVA, por sua vez causa necrose em batatas (HAIKONEN *et al.*, 2013). Além do envolvimento da HC-Pro (REVERS; GARCÍA, 2015), existem outros determinantes com mecanismos pouco claros, que participam do desenvolvimento dos sintomas, entre eles, por exemplo, a região C-terminal da P1 de plum pox virus (PPV) que contribui para sintomas suaves em plantas de *Nicotiana* e *Prunus persica* (MALIOGKA *et al.*, 2012). A P3 de soybean mosaic virus (SMV) está envolvida na necrose sistêmica em soja (HAJIMORAD; EGGENBERGER; HILL, 2005) e resposta HR (*Hypersensitive Response*) sistêmica em plantas de *Arabidopsis thaliana* (KIM *et al.*, 2010). A região N- e C-terminal de P3 contribui para sintomas severos em abobrinhas infectadas pelo vírus do mosaico amarelo da abobrinha (ZYMV) (DESBIEZ *et al.*, 2003) e para o mosaico clorótico em *Brassica juncea* infectado com o vírus do mosaico do nabo (*Turnip mosaic virus*-TuMV) (JENNER *et al.*, 2003). As regiões C-proximal de P3 e 6K1 de PPV estão envolvidos no desenvolvimento de manchas cloróticas em *N. benthamiana* (SÁENZ *et al.*, 2000), e sequências na junção P3/6K1 de PPV estão envolvidas no aparecimento de sintomas severos em *N. clevelandii* (RIECHMANN; CERVERA; GARCIA, 1995). Cui e Wang (2016) usaram dois mutantes de PPV (WT Δ 1CS/ Δ 1CS e WT Δ 2CS/ Δ 2CS) nos quais os primeiros e os últimos seis aminoácidos de 6K1 foram deletados a fim de evitar o reconhecimento pela NIa-Pro e a não clivagem nos locais de junção, resultando, portanto, nas proteínas fusionadas P3-6K1 ou 6K1-CI proteoliticamente não processáveis. Cui e Wang observaram que os mutantes perderam a capacidade de se replicar ou de mover-se intercelularmente. O mutante Δ 1CS de PPV quando comparado com o clone parental de PPV do tipo selvagem (WT) replicou-se em folhas de *N. benthamiana*, com um nível significativamente baixo e moveu-se muito devagar para as células vizinhas. A infecção sistêmica foi

significativamente atenuada, evidenciando que a eliminação dos sítios de clivagem entre P3 e 6K1 reduziu a infectividade do PPV. Portanto, as proteínas maduras P3 e 6K1 são importantes na manutenção da alta infectividade do PPV. A região da P3-6K1 não processada pode complementar parcialmente a função de 6K1 e P3, enquanto o 6K1-CI não processado não se sobrepõe funcionalmente com 6K1 e CI (CUI; WANG 2016). A clivagem na junção P3/6K1 leva à geração de três proteínas virais: P3, 6K1 e P3-6K1 (RIECHMANN; CERVERA; GARCIA, 1995). Portanto, a NIa-Pro cliva parcialmente na junção P3/6K1, gerando tanto a (poli) proteína P3-6K1 madura quanto a 6K1, ambos presentes durante a infecção por PPV (GARCÍA *et al.*, 1992; WALTERMANN; MAISS, 2006). Rodríguez-Cerezo e Shaw (1991) detectaram as proteínas P3 e P3-6K1 em plantas infectadas com o vírus do mosqueado da veia do tabaco (TVMV). A VPg associada a P3 de TEV estão envolvidos no desenvolvimento de sintomas fortes em *A. thaliana* (AGUDELO-ROMERO *et al.*, 2008). A região N-terminal da proteína CI de soybean mosaic virus (SBMV) também está envolvida em causar sintomas fortes na soja (ZHANG *et al.*, 2009). De acordo com o estudo de Martínez *et al.* (2016), a NIa-Pro pode interagir com 76 proteínas do hospedeiro envolvidas na resposta ao estresse, metabolismo e fotossíntese. Além disso, Faurez *et al.* (2012) observaram que a NIa-Pro de PVY da batata induziu necrose nas nervuras com sintomas de HR em tabaco, indicando que a NIa-Pro de potyvírus pode ser um determinante de patogenicidade. Por sua vez, Gong *et al.* (2020) usando um vetor de PVX como um sistema de expressão ectópica, descobriram que a NIa-Pro do pepper mottle virus (PepMoV) poderia induzir sintomas de mosaico graves em *N. benthamiana* e sugeriram que a NIa-Pro de PepMoV é um determinante da patogenicidade e um potente supressor de silenciamento transcricional no hospedeiro (silenciamento de genes transcricionais, *transcriptional gene silencing*).

Kim *et al.* (2015) estudaram a expressão transitória mediada por *Agrobacterium* de regiões codificantes de proteína de potyvírus em plantas de *Capsicum annuum* resistentes (gene *Pvr4*) e suscetíveis (gene *pvr4*) a potyvírus. Como resultado, os autores observaram resposta HR em folhas de pimenta que continham o gene *Pvr4* quando expressado a proteína NIb de PepMoV, de pepper severe mosaic virus (PepSMV) ou de PVY, demonstrando assim, que a NIb funciona como um fator de avirulência. Da mesma forma, Fellers *et al.* (2002) descobriram que a proteína NIb de PVY pode induzir uma resposta HR em plantas de tabaco com o gene *Rk*, que confere resistência ao nematoide *Meloidogyne incognita*. No entanto, observa-se que nem sempre o produto de um único gene é encontrado como o único determinante de respostas fitopatogênicas. Por exemplo, a região P1-HC-Pro é muitas vezes considerada responsável pelo efeito sinérgico dos potyvírus com outros vírus (REVERS; GARCÍA, 2015).

A maioria dos mecanismos de patogenicidade são incertos, mas pode-se inferir que os efeitos macroscópicos finais podem ser o produto da combinação da inferência aditiva de proteínas virais, implicando em alterações nos componentes e funções da maquinaria da fotossíntese, alterações em aspectos fisiológicos (ex. respiração, metabolismo de carboidratos), e alterações no espectro de produtos gênicos do hospedeiro, entre outras alterações (MAULE, 2001; VALLI; GARCÍA; LÓPEZ-MOYA, 2015; WANG, 2018).

A planta também pode induzir resposta de defesa por meio de necrose mediada por ácido salicílico e etileno regulada pela intensidade luminosa (KIM *et al.*, 2008). Alguns estudos revelaram que células infectadas por potyvírus apresentam alterações ultraestruturais, bioquímicas e metabólicas, tais como alterações nos cloroplastos, ou menos quantidade de cloroplastos e/ou tilacóides, além de alterações no acúmulo de amido e plastoglobulina (ZECHMANN; MÜLLER; ZELLNIG, 2003) em células adjacentes a tecidos necróticos (POMPE-NOVAK; WRISCHER; RAVNIKAR, 2001). Em outros casos, pode causar um efeito oposto, um aumento no número e tamanho da plastoglobulina nos cloroplastos, que se expressa na dilatação dos tilacóides e um menor teor de amido associado à alteração do metabolismo do cloroplasto (DÍAZ-VIVANCOS *et al.*, 2008; HERNÁNDEZ *et al.*, 2004). Por exemplo, o sunflower chlorotic mottle virus (SuCMoV) quando infecta o girassol causa uma diminuição na clorofila, na produção de proteínas e ROS (espécies reativas de oxigênio) apoplástico e também causa o aumento de açúcares solúveis, no teor de amido, na atividade de enzimas antioxidantes e a inibição da fotossíntese (ARIAS *et al.*, 2005; RODRÍGUEZ *et al.*, 2010, 2012). Vários genes do cloroplasto são regulados negativamente (*down-regulation*) da primeira à última fase no processo de desenvolvimento de sintomas (RODRÍGUEZ *et al.*, 2012).

4. Controle

Para o aparecimento da doença é necessária a presença simultânea do vírus, do vetor transmissor e de um hospedeiro vulnerável (REDINBAUGH; ZAMBRANO, 2014). A eficiência de controle só é obtida pela combinação de vários métodos de controle como práticas culturais, saneamento, resistência de plantas, abordagens biotecnológicas, entre outros (VALLI; GARCÍA; LÓPEZ-MOYA, 2015). Dado isto, diversas estratégias de manejo para o controle de fitoviroses devem ser consideradas, tais como o uso de pesticidas, que podem reduzir a população da colônia de insetos vetores (FERRO; MACKENZIE; MARGOLIES, 1980), embora não sejam muito eficazes na redução da incidência da doença para os vírus com modo de transmissão não-persistente (TOLER, 1985), e a aplicação de óleo mineral, que possivelmente atua interferindo na retenção do vírus nos estiletes dos afídeos (VALLI; GARCÍA; LÓPEZ-MOYA, 2015). Algumas das práticas culturais utilizadas são a alteração das épocas de

semeadura, plantio em locais com baixa população de vetores, o uso de plantas como barreira, o uso de cobertura morta que repele pulgões, e de coberturas de plástico absorvedor de UV (VALLI; GARCÍA; LÓPEZ-MOYA, 2015), bem como a erradicação de plantas alternativas que podem servir como reservatórios de vírus (STEWART *et al.*, 2014). O uso de sementes livres de vírus, estacas e/ou galhos indexados também devem ser considerados como parte da estratégia de manejo para prevenir reservatórios infectados. Dessa forma, é necessário implementar métodos precoces e sensíveis de diagnóstico ou detecção, bem como programas de certificação de sementes que fazem parte dos programas de controle e erradicação (VALLI; GARCÍA; LÓPEZ-MOYA, 2015).

Outro método que tem se mostrado eficiente na erradicação de vírus em diversas culturas de importância econômica é a crioterapia. Este método consiste no tratamento de um material vegetal infectado por um curto período em nitrogênio líquido para eliminar o vírus de dentro da planta utilizando a técnica de criopreservação (WANG; VALKONEN, 2008, 2009). Vários patógenos virais de diferentes grupos taxonômicos foram erradicados por esta técnica de crioterapia, como por exemplo o cucumber mosaic virus e banana streak virus (HELLIOT *et al.*, 2002), grape virus A (WANG *et al.*, 2003), plum pox virus em *Prunus* spp. (BRISON *et al.*, 1997), potato leaf roll virus em batata (WANG *et al.*, 2006), sweet potato feathery mottle virus e sweet potato chlorotic stunt virus (WANG; VALKONEN, 2008). Ita *et al.* (2020) demonstraram a erradicação de YMV em *D. rotundata* por crioterapia, em que botões criotratados produziram mudas saudáveis após propagação *in vitro*. Em acessos de *D. alata*, *D. cayennensis-rotundata* e *D. trifida* do BRC-TP, em um ciclo de saneamento (termoterapia e cultura do meristema), pesquisadores obtiveram eficiência moderada na eliminação de vírus, com 9,7% dos acessos completamente livres de YMV, YMMV, badnavírus, maculavírus e potexvírus (UMBER *et al.*, 2020). Esta estratégia de controle é mais cara em comparação com a crioterapia (UMBER *et al.*, 2020; WANG; VALKONEN, 2008), e recentemente o tratamento com ozônio dissolvido em água está sendo explorado como uma alternativa mais viável, sendo possível eliminar potyvírus em até 63,5% (GONZÁLEZ RAMÍREZ *et al.*, 2020).

A proteção cruzada é uma alternativa promissora para o controle de doenças virais, no qual plantas inoculadas com um vírus que causam sintomas leves manifestam resistência à infecção subsequente devido a um isolado mais severo do mesmo vírus inoculado ou de outro intimamente relacionado (CONG *et al.*, 2019; FULTON, 1986; KURIHARA; WATANABE, 2003). Esta estratégia de manejo tem sido usada para controlar alguns potyvírus, como o bean yellow mosaic virus, clover yellow vein virus (NAKAZONO-NAGAOKA *et al.*, 2009), watermelon mosaic virus (KOSAKA; FUKUNISHI, 1997), papaya ringspot virus (YEH, 1984;

YOU *et al.*, 2005), e o soybean mosaic virus (KOSAKA Y, 1993), entre outros. Além disso, a proteção cruzada pode ser ativada pela introdução de uma sequência antisense similar aos genes do vírus, como a CP, replicase, e a proteína de movimento, que são multifuncionais (LIU *et al.*, 2009; MURRY *et al.*, 1993). A proteção mediada por CP permitiu demonstrar que sequências virais que codificam proteínas estruturais e não estruturais, bem como regiões não codificantes conferem resistência ao vírus (AZAD; AMIN; SIDIK, 2014; PRINS *et al.*, 2008). Como exemplos temos o desenvolvimento de linhagens de milho modificadas com resistência genética ao maize dwarf mosaic virus (MDMV) (KANNAN; ISMAIL; BUNAWAN, 218), a resistência ao potyvírus mediada por CP em batata resistente a PVY e a abobrinha resistente ao ZYMV (VALLI; GARCÍA; LÓPEZ-MOYA, 2015).

Usar plantas resistentes é uma das estratégias mais eficazes e relativamente mais acessíveis economicamente, especialmente por ser uma abordagem útil que fornece resistência estável aos potyvírus. Não obstante, há preocupação no aparecimento de isolados que quebrem a resistência e durabilidade (HISLOP *et al.*, 2021). No caso de haver bancos de germoplasma, o passo inicial é identificar a fonte de resistência por meio da triagem de germoplasma coletado (REDINBAUGH; ZAMBRANO, 2014). A triagem bem-sucedida continua com a identificação de marcadores ligados a genes ou QTL (loci que influenciam características quantitativas, *Quantitative Trait Loci*) que exercem resistência (SCOTT; NELSON, 1972), marcadores morfológicos (SCOTT, 1989) e análise de polimorfismos (MCMULLEN, 1991).

Cada vez contamos com mais ferramentas no manejo de doenças virais, e a engenharia genética nos permite incorporar características de resistência a vírus às plantas de interesse. No início, os transgênicos resistentes a vírus baseavam-se na resistência derivada do patógeno (AZAD; AMIN; SIDIK, 2014; PRINS *et al.*, 2008). Este conceito de resistência derivado de patógenos foi utilizado para o desenvolvimento do mamão transgênico, utilizando um mecanismo de silenciamento de RNA mediado por CP para o tratamento eficaz da doença ocasionada pela espécie *Papaya ringspot virus* (AZAD; AMIN; SIDIK, 2014; SOUZA JÚNIOR; NICKEL; GONSALVES, 2005). Este mecanismo de silenciamento gênico pós-transcricional é uma tecnologia promissora para o tratamento da doença como por exemplo às causadas por PRSV (AZAD; AMIN; SIDIK, 2014; SOUZA JÚNIOR; NICKEL; GONSALVES, 2005). A planta possui uma resistência natural ao vírus devido ao silenciamento do RNA (AZAD; AMIN; SIDIK, 2014; PRINS *et al.*, 2008). O silenciamento do RNA está sendo explorado para obtenção de resistência através do desenho de transgênicos específicos que expressam estruturas de *hairpins* (construções invertidas) ou mesmo através da aplicação direta de moléculas de dsRNA (BEHBOODIAN *et al.*, 2012; FAHIM *et al.*, 2010; PRAVEEN *et al.*,

2010). A proteção cruzada, também está relacionada com o silenciamento do RNA, onde RNAs com estruturas invertidas auto-complementares podem silenciar genes virais (CHEN *et al.*, 2004; HILY *et al.*, 2007).

Conforme indicado nos parágrafos anteriores, não existe um método único de controle de vírus. Um controle eficiente requer conhecimento prévio do agente causador, do hospedeiro, das condições ambientais e das técnicas ou métodos disponíveis que se enquadram na realidade de cada um.

5. Sequenciamento de alto desempenho (*High-Throughput Sequencing*, HTS)

As metodologias de sequenciamento de DNA evoluíram muito rapidamente desde o sequenciamento de Sanger (SANGER; COULSON, 1975), tanto na capacidade de sequenciamento quanto em suas aplicações. Em 2001 o sequenciamento do genoma humano foi concluído (LANDER *et al.*, 2001; VENTER *et al.*, 2001). Após a conclusão do sequenciamento do genoma humano, muitas empresas foram criadas para desenvolver novas tecnologias de “sequenciamento de próxima geração” (*Next-Generation Sequencing*, NGS) e, como consequência, a bioinformática tornou-se uma disciplina científica muito relevante para diversas áreas da ciência (LEVY; BOONE, 2019). Por cerca de uma década, o sequenciamento de “próxima geração” tem sido usado para se referir a métodos de sequenciamento de alto rendimento que nos permitem realizar milhões a bilhões de observações em uma única execução (LEVY; BOONE, 2019). Consequentemente, em menos de 15 anos, os custos do sequenciamento diminuíram (LEVY; BOONE, 2019). Os equipamentos para o HTS tornaram-se acessíveis em instituições de pesquisa e até mesmo em laboratórios individuais. Portanto, a diminuição de custos e aumento da acessibilidade permitiram que pesquisadores utilizassem tecnologias de HTS em diversas áreas da ciência (REUTER; SPACEK; SNYDER, 2015). As tecnologias de HTS permitem o sequenciamento simultâneo de amostras, e assim, a detecção de qualquer organismo ou agentes infecciosos presentes nas mesmas (OLMOS, A. *et al.*, 2018).

Diferentes tipos de extrações de ácidos nucleicos, tais como de ssRNA, RNA total, dsRNA, *small* RNA e preparações enriquecidas com partículas virais (VLPs) podem ser explorados na tecnologia de HTS para estudar viroma de vírus de RNA (ROOSSINCK; MARTIN.; ROUMAGNAC, 2015). Caso o vírus tenha um genoma de DNA circular (geminivírus ou nanovírus, ssDNA) ou pseudocircular (caulimovírus, dsDNA), o enriquecimento viral pode ser realizado com a ajuda da técnica de RCA (amplificação em círculo rolante, *rolling circle amplification*) (IDRIS *et al.*, 2014; INOUE-NAGATA *et al.*, 2004; ROSÁRIO *et al.*, 2013).

5.1. Plataformas HTS

As plataformas HTS compartilham três etapas: (i) fragmentação de DNA para criar a biblioteca, (ii) adição de adaptadores de DNA sintético a cada fragmento e (iii) sequenciamento de cada fragmento (VILLAMOR *et al.*, 2019). No caso de RNA (análise do transcriptoma), a seleção do tipo de RNA é realizada geralmente primeiro por: (i) depleção de RNA ribossômico e (ii) enriquecimento de RNAs poliadenilados (ADAMS; FOX, 2016; GAAFAR *et al.*, 2021; VILLAMOR *et al.*, 2019). Após o procedimento de seleção, o RNA fragmentado é submetido a uma reação de transcrição reversa e uma biblioteca de cDNA é criada (VILLAMOR *et al.*, 2019). As plataformas de HTS da Illumina, PacBio, SOLiD (detecção de fluorescência) e 454 (pirosequenciamento) usam o método de sequenciamento de nucleotídeos por meio de detecção óptica, enquanto o Ion Torrent utiliza a detecção não óptica, pois registra mudanças de pH durante a polimerização. Nas plataformas de HTS da Illumina, Ion Torrent, Roche 454 e SOLiD, que usam a fonte proximal de nucleotídeos, cada sequência lida é derivada de DNA amplificado clonalmente, enquanto as sequências que provém de PacBio e Oxford Nanopore são originárias de uma única molécula (VILLAMOR *et al.*, 2019). Estas plataformas de sequenciamento químico (LEVY; MYERS, 2016) da Illumina, Ion Torrent, PacBio e Roche 454 usam a polimerase para conduzir o sequenciamento por reação de síntese, já a SOLiD usa sequenciamento de ligação e a Oxford Nanopore determina diretamente a sequência de DNA conforme passa por um nanoporo. Uma lista de tipos e características dos sequenciadores massivos principais é detalhada na Tabela 5 (HERNÁNDEZ *et al.*, 2020).

Tabela 5 - Tipos e características dos principais sequenciadores massivos.

Sequenciador	Companhia	Ano ^a	Química	Comprimento máximo de leitura	Máx. desempenho / carreira	Taxa de erro (%)	Máx. n° sequências / carreira
454 GS FLX	Roche	2004	Pirosequenciamento-Emulsão de	700-800 pb	0,7 Gb	1	5×10^5
Analizador de genoma GA	Solexa	2006	Synthesis-terminators-clusters	2×100 pb	1 Gb	~0,1	1×10^8
SOLiD®	Life Technologies	2007	Ligação-emulsão	75 + 35 pb	150 Gb	~5	1.400×10^6
Ion Torrent PGM™	Life Technologies	2010	síntese-emulsão-deteccão de prótons	200-400 pb	2 Gb	~1	$0,4-5,5 \times 10^6$
Proton™	Life Technologies	2012	síntese-emulsão-deteccão de prótons	200 pb	10 Gb (atualmente ~100 Gb)	~1	$60-80 \times 10^6$
PacBio®	Pacific Biosciences	2010	SMRT	6-8 kb (máx. 15 kb)	3 Gb / dia	~13	370.000×10^6
HiSeq	Illumina	2010	Síntese-terminadores-clusters	2×150 pb (250 pb em modo <i>rapid run</i>)	1.500 Gb	~0,1	5.000×10^6
MiSeq	Illumina	2011	Síntese-terminadores-clusters	2×300 pb	15 Gb	~0,1	25×10^6
Nextseq	Illumina	2014	Síntese-terminadores-clusters	2×150 pb	120 Gb	~0,1	260×10^6

SMRT: single molecule real-time

^a Ano de introdução no mercado.Extraído de: (HERNÁNDEZ *et al.*, 2020).

A primeira plataforma de HTS foi lançada em 2005 como 454 e posteriormente foi adquirida pela Roche em 2007 quando passou a se chamar Roche 454. Nesta plataforma uma molécula molde é capturada por “beads”, e então é carregada em um poço de placa de picotitulação para sua amplificação por PCR em emulsão, para finalmente ser sequenciada no pirosequenciador (ROTHBERG; LEAMON, 2008). A tecnologia Illumina foi lançada em 2005 pela Solexa (hoje Illumina), sendo baseada no sequenciamento por síntese que utiliza terminadores marcados com fluorescência. Esse processo envolve a amplificação clonal de fragmentos de DNA ligados à superfície da câmara de reação (a célula de fluxo, *the flowcell*) (BENTLEY *et al.*, 2008). O sistema de sequenciamento NovaSeq 6000 da Illumina (MODI *et al.*, 2021) pode fornecer até 20 bilhões de leituras por corrida com um comprimento máximo de leitura emparelhada de 150 bp. A plataforma SOLiD™, lançada em 2007 pela empresa *Applied Biosystem*, utiliza o método de ligação por DNA ligase. No processo de sequenciamento algumas etapas são seguidas: ciclos repetitivos de ligação de sonda marcada com fluorescência, a lavagem de sondas não ligadas e captura de imagens (VALOUEV *et al.*, 2008), que determina as sequências de nucleotídeos com duas leituras de cada base. Uma das versões desta plataforma é o sistema 5500xl W, que foi lançado em 2013, o qual fornece até 6 bilhões de leituras por execução. A tecnologia *Ion Torrent* possui semelhanças com a plataforma de pirosequenciamento Roche 454. A reação de síntese é realizada em cada micropoço, onde é determinada a sequência. Esta determinação das bases é realizada medindo-se a mudança de pH, que resulta na liberação de íons de hidrogênio durante a incorporação das bases (ROTHBERG *et al.*, 2011). Os instrumentos Ion PGM 314, 316, 318 e S5 da *Ion Torrent* podem gerar tamanhos de leitura de até 400 pb, e o Proton 1 pode produzir até 80 milhões de leituras (menos que os sistemas Illumina e SOLiD™). O sistema *PacBio - Pacific BioSciences* foi lançado em 2010, nas quais são sequenciadas moléculas únicas usando adaptadores com estruturas de *hairpin* para um molde de ssDNA fechado, chamado de SMRTbell (molécula única, sequenciamento em tempo real; *Single Molecule Real Time sequencing*) (RHOADS; AU, 2015). O molde SMRTbell é colocado em uma câmara de tamanho em zeptolitro (10^{-21} L), e uma guia de onda em modo zero é ligada a uma única molécula de polimerase, o qual vai para o fundo da câmara. Em seguida é adicionado nucleotídeos marcados com fluorescência no grupo fosfato, para serem detectados em tempo real. O instrumento *Sequel* pode entregar até 370.000 leituras com comprimentos de até 20 kb, porém possui uma alta taxa de erro de incorporação de bases (VILLAMOR *et al.*, 2019). Por último, a tecnologia *Oxford Nanopore*, por meio de sua plataforma *nanopore*, sequencia uma única molécula. Nessa tecnologia, as mudanças características na corrente induzida são medidas quando as bases passam por um nanoporo biológico, ancorado a ele uma

proteína motora molecular (LU; GIORDANO; NING, 2016). A taxa de erro desta plataforma é maior que a do *PacBio* e o desempenho é inferior ao da Illumina. No entanto, o sistema MinION da *Oxford Nanopore* vem sendo distribuído desde 2014 e fornece resultados rentáveis em tempo real quando não é requerido grande quantidade de conjuntos de dados (LU; GIORDANO; NING, 2016). Além disso, o sistema fornece leituras ultra-longas de até 200 kpb, é portátil e o sequenciamento pode ser realizado em tempo real (MINEI; HOSHINA; OGURA, 2018).

5.2. Aplicações de HTS na detecção de vírus de plantas

As aplicações do HTS são diversas, como por exemplo, em estudos genômicos de populações virais (GONZÁLEZ-CANDELAS; PATIÑO-GALINDO; VALIENTE-MULLOR, 2018), em estudos de biologia evolutiva viral (ARENAS et al., 2018), em análises filogenéticas e filodinâmicas (BROMHAM et al., 2018), na metagenômica (ZHANG et al., 2019), entre outras. O uso do HTS na detecção e descoberta de vírus de plantas teve seu início em 2006, onde as primeiras aplicações foram para estudos de biodiversidade e ecologia de fitovírus (WREN et al., 2006). Em 2008, Melcher et al. detectaram vírus pertencentes a duas famílias (*Caulimoviridae* e *Flexiviridae*) a partir de amostras de *Ambrosia psilostachya* DC. Em 2009, a tecnologia foi utilizada para sequenciar o genoma de um membro do gênero Tymovirus que infectou Tallgrass prairie (MUTHUKUMAR et al., 2009). Kreuze et al. (2009) extraíram pequenos RNAs de plantas de batata-doce e através do uso da plataforma Illumina (1,2 milhões de leituras “reads”, 1-28 nt) montaram sequências genômicas completas de sweetpotato feathery mottle virus (SPFMV; família *Potyviridae*) e sweetpotato chlorotic stunt virus (SPCSV; família *Closteroviridae*) (KREUZE et al., 2009). Da mesma forma, em 2009, utilizando as plataformas Roche 454, outras sequências virais foram identificadas em plantas cultivadas para estudos etiológicos de doenças virais (ADAMS et al., 2009; DONAIRE et al., 2009). Nos últimos anos, o HTS tem sido usado para a detecção rápida e confiável de sequências virais de plantas em locais geográficos onde o fitopatógeno não deveria estar presente ou são desconhecidos (às vezes considerados quarentenários). A exemplo tem-se o estudo realizado por Jarugula et al. (2020) que por meio da plataforma Illumina HiSeq2500 (Novaseq 6000) relataram pela primeira vez a presença de grapevine red globe virus (GRGV; gênero *Maculavirus*, família *Tymoviridae*) em campos de produção de videiras em Washington, EUA (JARUGULA et al., 2020).

Concluindo, o aumento do uso de HTS tem levado a descoberta de novas sequências virais tanto em plantas cultivadas como não cultivadas (BARBA; CZOSNEK; HADIDI, 2013; DIAZ-LARA et al., 2019; KREUZE et al., 2009; MALIOGKA et al., 2018; MUTHUKUMAR et al., 2009; NAVARRO et al., 2018; POOGGIN, 2018; RWAHNIH et al., 2012; RWAHNIH

et al., 2015). Por fim, a tecnologia está revolucionando o conhecimento em relação a compreensão da natureza ubíqua dos vírus e permite o descobrimento de uma grande diversidade viral não caracterizada anteriormente (SIMMONDS *et al.*, 2017; YUTIN *et al.*, 2018).

5.3. Análise de dados HTS

Armazenamento dos dados brutos: um grande número de leituras requer etapas complexas de pré-processamento antes da obtenção das informações. É necessário um computador de alto desempenho que execute algoritmos compilados em pacotes de softwares de bioinformática unificados. No conjunto de dados de HTS armazenado no formato FASTQ encontra-se as sequências para cada leitura.

Qualidade: o controle de qualidade é baseado em pontuações de qualidade Phred (Q), comumente usando o programa FASTQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), que pode ser usado em sistemas Windows, Mac e Linux, e também está disponível na interface web do Galaxy (<https://usegalaxy.org/>). O software FASTQC fornece informações sobre a qualidade da sequência de uma biblioteca e auxilia na escolha dos melhores parâmetros de recorte para evitar perdas de informações das sequências (BLAWID; SILVA; NAGATA, 2017).

Trimming: a trimagem das sequências visa melhorar a qualidade das leituras, e é uma etapa muito importante em todo processamento de dados (MACMANES, 2014). A trimagem excessiva pode gerar perda de informações e a falta de trimagem pode gerar dados com contaminações (*primers*, adaptadores, códigos de barras, entre outros), afetando as análises subsequentes dos dados. A maioria dos métodos disponíveis para trimagem são baseados na trimagem de adaptadores, outros podem separar leituras multiplexadas com base em códigos de barras. Um dos programas de trimagem mais utilizados atualmente é o *Trimmomatic*, que executa uma variedade de tarefas úteis de ‘trimming’ (processo de corte) para dados únicos e emparelhados de Illumina (BOLGER; LOHSE; USADEL, 2014). O algoritmo é habilitado para cortar sequências de adaptadores e/ou outras sequências específicas (ILLUMINACLIP), sequências que estão abaixo do limite de qualidade definido pelo usuário através de uma janela deslizante (SLIDINGWINDOW), bases do início de uma leitura quando estiver abaixo de um limite de qualidade (LEADING), bases do final de uma leitura quando abaixo de um limite de qualidade (TRAILING), cortar leituras em um comprimento especificado (CROP), cortar um número especificado de bases desde o início da leitura (HEADCROP), e leituras se estiverem abaixo de um comprimento especificado (MINLEN), entre outros (BOLGER; LOHSE; USADEL, 2014).

Montagem de contigs: após a trimagem, as sequências são montadas em *contigs* (sequências de leituras montadas). Uma das estratégias utilizadas para a obtenção de um genoma viral é através da montagem do tipo *de novo* (*de novo strategy*), que não é necessário o conhecimento prévio das sequências genômicas. As etapas das estratégias básicas para a montagem *de novo* consistem na montagem dos *contigs*, *scaffolding* (andaime, sequências de *contigs* montadas) e preenchimento das lacunas (*gaps*) (MILLER; KOREN; SUTTON, 2010; NAGARAJAN; POP, 2013; PASZKIEWICZ; STUDHOLME, 2010; SIMPSON; POP, 2015). As leituras são montadas para formar uma sequência consenso denominada *contig* com ausência de *gaps* (SOHN; NAM, 2018), enquanto na etapa de *scaffolding*, os *contigs* são conectados por grandes insertos. Uma vez estruturados os *contigs*, os *gaps* permanecem quando não há sobreposição entre os *contigs*. As bases indefinidas e as distâncias aproximadas são estimadas a partir das grandes leituras de inserção e os *gaps* podem ser cuidadosamente preenchidos para completar a montagem (BOETZER; PIROVANO, 2012; KOSUGI; HIRAKAWA; TABATA, 2015; LUO *et al.*, 2015; PAULINO *et al.*, 2015; TSAI; OTTO; BERRIMAN, 2010). A utilização do gráfico de *Bruijn* é a abordagem mais conhecida para montar leituras curtas (MILLER; KOREN; SUTTON, 2010; NAGARAJAN; POP, 2013; SIMPSON; POP, 2015), que é empregado por vários montadores. Os algoritmos baseados em *de Bruijn* são baseados em alinhamentos de pares que detectam sobreposições e são considerados computacionalmente eficientes, pois resolvem problemas difíceis de montagem. Esse algoritmo divide as leituras em subcadeias de tamanho fixo “*k*” chamadas de *k*-mers (COMPEAU; PEVZNER; TESLER, 2011).

Vários programas de montagem são usados para leituras curtas de dados metagenômicos, os quais podem ser compilados em plataformas de interface web ou instalados em diferentes sistemas operacionais. Dentre os montadores usados de genomas virais de plantas e que usam o algoritmo de *Bruijn*, podem-se destacar os seguintes:

SOAPdenovo2 (LUO *et al.*, 2012) – suporta *k*-mer escasso (63/127). Seu ponto forte é que cada etapa (correção de erros, montagem de *contigs* ou *scaffolding*) da montagem *de novo* pode ser realizada separadamente de acordo com a necessidade do usuário. Além disso, os novos módulos foram desenhados para resolver regiões mais longas de repetição na montagem de *contigs*, aumentar o tamanho da montagem e a cobertura no *scaffold*, e melhorar o fechamento dos *gaps*.

Velvet (ZERBINO; BIRNEY, 2008) – utiliza o gráfico de *Bruijn* bidirecionalmente. Possíveis erros de montagem são corrigidos pelos chamados algoritmos “tour-bus” ao nível do gráfico de montagem (remove nós abaixo de um limite de cobertura/comprimento). O programa foi desenvolvido especificamente para leituras muito curtas e informações finais emparelhadas.

MetaVelvet (NAMIKI *et al.*, 2012) e MetaVelvet-SL (AFIAHAYATI; SATO; SAKAKIBARA, 2015) – são extensões do Velvet de montagem de genoma único. Os procedimentos do Meta Velvet consistem na (i) construção do gráfico de *Bruijn* a partir de leituras de sequências de espécies mistas (primeiro estágio Velvet); (ii) detecção de múltiplos picos da frequência de *k*-mers; (iii) decomposição do gráfico de *Bruijn* por conectividade, onde o grafo de *Bruijn* inicial se decompõe em subgrafos conectados e ocorre a decomposição por valores de cobertura dos nós; e (iv) montagem de *contigs* e *scaffolds*. As montagens são de maior qualidade do que as do Velvet e do SOAPdenovo quando aplicadas a leituras de sequências de metagenômica. O MetaVelvet tem baixa precisão e sensibilidade na detecção de nós quiméricos no gráfico de *Bruijn*, o que impede a geração de *contigs* e *scaffolds* mais longos. Já a ferramenta MetaVelvet-SL supera o MetaVelvet original por meio de automação supervisionada que permite uma classificação dos nós quiméricos, melhorando assim o desempenho e reconstruindo montagens mais longas e precisas.

IDBA-UD (PENG *et al.*, 2012) – projetado para dados de sequenciamento de células únicas ou sequenciamento metagenômico com coberturas de sequências desiguais. O programa usa vários limites relativos para eliminar *k*-mers errôneos em regiões de alta e baixa cobertura. Utiliza a montagem local com informações de extremidade emparelhadas para resolver o problema de ramificação através da remoção de nós curtos de baixa cobertura com base em nós vizinhos. Na correção de erros, o programa corrige leituras para regiões de alta cobertura que podem ser alinhadas com *contigs* de alta confiabilidade.

SPAdes (BANKEVICH *et al.*, 2012) – foi projetado para sequenciamento de célula única. A versão atual do SPAdes (v. 3.12.0) funciona com leituras do tipo *Illumina* ou *IonTorrent* e é capaz de fornecer montagens híbridas usando leituras de *PacBio*, *Oxford Nanopore* e Sanger. Oferece suporte a leituras do tipo “*paired-end*”, “*mate-pairs*” e “*unpaired*” (BANKEVICH *et al.*, 2012; MEDVEDEV *et al.*, 2011). O SPAdes constrói o gráfico de *Bruijn* a partir de leituras curtas. Este gráfico passa por um processo de remoção de arestas erradas os quais podem ser gerados por erros no sequenciamento ou no equipamento. Depois da simplificação do gráfico de montagem, o SPAdes mapeia leituras curtas e longas, e volta ao gráfico de montagem usando as informações de alinhamento para executar a resolução repetida e o *scaffolding*, que visa a construção de caminhos corretos e contínuos para o que está sendo montado (PRJIBELSKI *et al.*, 2020). Este programa é usado em dados de cobertura desigual e, atualmente, suporta leituras longas emparelhadas de *Illumina* (2 x 150 e 2 x 250) e leituras longas de *PacBio* (montagem híbrida). Também inclui um corretor de erros e a montagem é interativa com vários valores *k*.

O ABySS (*Assembly By Short Sequences*) (SIMPSON *et al.*, 2009) – é um programa computacionalmente eficiente, que usa a tabela "hash" de *k*-mers distribuída em um sistema de agrupamentos para reduzir o requisito de memória RAM do computador. No desenvolvimento do ABySS, um processo semelhante ao Velvet foi desenvolvido para lidar com erros de leitura. Primeiro, o algoritmo remove ramificações curtas. Erros mais complexos que formam bolhas divergentes no gráfico são removidos para o registro em uma etapa posterior. Essas duas etapas de depuração são repetidas para corrigir erros de leitura próximos. Este programa paraleliza esses processos para permitir que o montador atue com eficiência na montagem dos genomas de tamanho desejado. O ABySS pode paralelizar a montagem de bilhões de leituras curtas em um *cluster* de hardware comum, permitindo assim uma memória disponível para o processo de montagem para lidar com genomas de qualquer tamanho.

MEGAHIT (LI *et al.*, 2015) – é um montador de dados metagenômicos grandes e complexos. Este programa reúne todos os dados como um todo, ou seja, não precisa de nenhum processamento prévio como a partição e a normalização. Por este motivo consegue realizar as montagens com pouca memória, é muito rápido e é interativo permitindo vários valores de *k*.

Vários trabalhos foram realizados na comparação de montadores de dados de HTS. Por exemplo, Barrero *et al.* (2017) sequenciaram pequenos RNAs (21-25 nt) isolados de 12 plantas, e as leituras do sequenciamento foram montadas usando os programas Velvet (ZERBINO; BIRNEY, 2008), SPAdes (BANKEVICH *et al.*, 2012) e SOAPdenovo (LUO *et al.*, 2015), utilizando *k*-mers individuais de 15 (k15), 19 (k19) e 21 (k21), bem como conjuntos de *k*-mers de 15,17, 19 (k15-17-19) e 15, 17, 19 e 21 (k15-17-19-21). Como resultado, Barrero *et al.* observaram que o SPAdes k15-17-19 e SPAdes k15-17-19-21 produziram sequências mais longas no processo de scaffolding e o SPAdes k15-17-19 produziu um maior número de bases montadas (o melhor resultado). Barrero *et al.* atribuem o melhor resultado ao fato de SPAdes usar a abordagem única para construir progressivamente a montagem de *contigs*, em que o primeiro usa o *k*-mer mais curto na primeira corrida e depois, com base nesta corrida, continua montando as sequências usando *k*-mers de maiores tamanhos. Este recurso iterativo não está disponível no Velvet, ABySS e SAOPdenovo. Por sua vez, Lai *et al.* (2015) compararam os montadores ABySS, EDBA-UD, CABOG (MILLER *et al.*, 2008), MetaVelvet e SOAPdenovo com dados metagenômicos por meio de diversos parâmetros como a taxa de cobertura, erros de montagem e comprimentos de *contigs*. Os autores concluíram que esses programas eram complementares entre si na montagem de sequências metagenômicas.

6. Filogenia molecular

A filogenética se preocupa com a reconstrução da história evolutiva (árvore filogenética) de espécies, indivíduos ou genes relacionados a partir de dados genéticos por alinhamentos múltiplos de suas sequências (BROMHAM *et al.*, 2018).

Através da filogenética foram desenvolvidos métodos para inferir filogenias e estimar datas de divergência de um conjunto de dados. É possível que com taxas (*rate*) de evolução constante sejam fáceis de inferir datas de divergência (DRUMMOND *et al.*, 2006). No entanto, na prática existe uma variação das taxas de evolução entre linhagens, representando um problema na inferência filogenética. Portanto, o uso de modelos que assumem taxas independentes de evolução molecular em cada ramo ajuda a evitar erros na estimativa de datas de divergência e análises filogenéticas (HO; JERMIIN, 2004). Sob tal suposição de taxas independentes pode-se inferir filogenia, mas não estimar taxas moleculares ou tempos de divergência, porque as contribuições individuais de taxa e tempo para a evolução molecular não podem ser separadas. Se a taxa evolutiva e o tempo ao longo de cada ramo só puderem ser estimados como seu produto, então a posição da raiz da árvore não pode ser estimada sem suposições adicionais, ou seja, sem a adição de *outgroups* (usada apenas para enraizamento; caso contrário, a árvore é simétrica) ou de um processo de substituição não-reversível (*non-reversible substitution process*) (DRUMMOND *et al.*, 2006; HASTINGS, 1970; METROPOLIS *et al.* 1953; YAP; SPEED, 2005). Felsenstein (1981) sugeriu pela primeira vez a alternativa desenraizada (*unrooted*) ao relógio molecular (FELSENSTEIN, 1981). As árvores sob o modelo *unrooted* seriam um tipo equivalente de árvore enraizada, desde que a raiz da árvore pudesse ser localizada em qualquer lugar ou ponto de ramificação da árvore sem afetar a probabilidade. Assim, a raiz é como uma polia, onde todas as partes da árvore de um lado da raiz são movidas para baixo e todas as partes do outro lado são movidas para cima na mesma proporção, a probabilidade permanece inalterada (princípio da polia). As análises de dados genéticos por meio de algoritmos e métodos computacionais viáveis descritas por Felsenstein (FELSENSTEIN, 1981) visaram calcular a probabilidade (probabilidade filogenética da árvore, verossimilhança filogenética) dos alinhamentos de sequências de ácido nucléico e assim gerar árvores evolutivas aplicando a técnica de máxima verossimilhança (*maximum-likelihood estimation*- MLE). Com o surgimento da inferência Bayesiana (HUELSENBECK; RONQUIST, 2001), algoritmos denominados “Metropolis - Hastings” foram utilizados com modelos complexos e com a possibilidade de integrar várias fontes de dados para integrá-los em uma única análise (BROMHAM *et al.*, 2018).

A inferência Bayesiana realiza reconstruções filogenéticas, avalia incertezas nas reconstruções e faz inferências estatísticas por meio do amostrador de Cadeias de Markov via Monte Carlo (*Markov Chain Monte Carlo*, MCMC) (LARGET; SIMON, 1999), que faz amostragens das árvores e modela valores dos parâmetros usados a partir da distribuição posterior. As inferências estatísticas baseiam-se na amostragem, que fornecerá a árvore mais provável, com suas respectivas probabilidades posteriores para cada clado (LARGET; SIMON, 1999). As informações da árvore mais provável são qualitativamente semelhantes à de uma árvore de máxima verossimilhança com proporções de “*Bootstrap*” e permite inferências adicionais sobre a topologia da árvore, comprimentos dos ramos e os valores dos parâmetros do modelo (LARGET; SIMON, 1999). A vantagem da inferência bayesiana é a facilidade de interpretação dos resultados, a possibilidade de incorporação de informações prévias, além de vantagens computacionais (HUELSENBECK; RONQUIST, 2001). Por exemplo, segundo Larget e Simon (1999) para obter-se a melhor topologia de uma árvore com 31 táxons com uma probabilidade posterior de 0,413 e erro padrão estimado de Monte Carlos de 0,013 foram necessárias sete horas de tempo de CPU (PC Pentium II 300 MHz rodando sob Solaris x 86), e para analisar os mesmos dados com o mesmo grau de precisão com o método de *bootstrapping* foram necessárias 1400 réplicas de *bootstrap* e 175 dias de tempo de CPU (LARGET; SIMON, 1999).

Para a inferência bayesiana, a distribuição de probabilidade anterior (*prior probability*) deve ser especificada na forma da filogenia (idades dos nós e ordem de ramificação) (DRUMMOND *et al.*, 2006). Ou seja, o teorema de Bayes estabelece a distribuição de probabilidade posterior dada pelo produto da distribuição anterior e a verossimilhança (*likelihood*), dividido pela probabilidade dos dados, que é representada pela seguinte fórmula: $P(\theta | D) = P(\theta) P(D | \theta) / P(D)$, onde, $P(\theta | D)$ é a distribuição de probabilidade dos parâmetros condicional aos dados obtidos (probabilidade posterior, *posterior probability*), $P(\theta)$ é a distribuição de probabilidade dos parâmetros (probabilidade anterior, *prior probability*), $P(D | \theta)$ é a probabilidade dos dados condicional aos parâmetros obtidos (verossimilhança, *likelihood*) e $P(D)$ é a probabilidade dos dados (BROMHAM *et al.*, 2018). A pontuação de verossimilhança é avaliada usando o modelo de verossimilhança, que inclui a árvore e o modelo de substituição (BROMHAM *et al.*, 2018). O “*prior*” representa a crença sobre os parâmetros do modelo de verossimilhança, sem considerar nenhuma informação dos dados (BROMHAM *et al.*, 2018). Em alguns casos, a distribuição dos *priors* pode ser gerada usando modelos de processos evolutivos. Esses modelos podem ter parâmetros chamados hiperparâmetros e suas distribuições anteriores são chamadas de *hiperpriors*. O modelo de verossimilhança e os *priors*

são chamados coletivamente de "modelo hierárquico" porque envolvem várias camadas de modelos (BROMHAM *et al.*, 2018).

A análise Bayesiana em plataformas avançadas como o software BEAST (BOUCKAERT *et al.*, 2014, 2019) requer a especificação de um modelo de substituição, um modelo de relógio e um modelo de população dinâmico, com o qual geramos a estrutura filogenética de uma árvore, uma rede filogenética ou uma combinação dos dois (BOUCKAERT *et al.*, 2019). Esses modelos induzem distribuições de probabilidade usando a abordagem *Markov Chain Monte Carlo* (MCMC) (BOUCKAERT *et al.*, 2019). O algoritmo MCMC tem permitido o uso de inferência Bayesiana em várias áreas da vida humana, tais como: nas ciências comportamentais, nas finanças, na saúde humana, no controle de processos, na avaliação de risco ecológico, na avaliação de risco de sistema de engenharia, entre outros (KELLY; SMITH, 2009). O método MCMC pode incluir modelos de várias categorias: paramétrica, semiparamétrica, *frailty* e outros métodos não tradicionais (LIN, 2014). Por sua vez, os modelos paramétricos Bayesianos oferecem opções comuns que incluem os modelos *Weibull*, valor extremo, log-normal e gamma (LIN; ASPLUNDA; PARIDAA, 2014). Nos modelos semi-paramétricos uma estratégia de modelagem mais geral é oferecida, com menos suposições, onde as taxas de erro são descritas semi parametricamente ou os “*priors*” são desenvolvidos por meio de um processo estocástico (LIN; ASPLUND, 2015). Os métodos de amostragem baseados em Monte Carlo (MC) foram desenvolvidos para avaliar integrais posteriores de alta dimensão, sendo os mais comuns: (i) amostragem *Metropolis-Hastings*, que é um algoritmo de amostragem MCMC bem conhecido, que foi desenvolvido pela primeira vez por Metropolis *et al.* (METROPOLIS *et al.*, 1953) e então generalizado por Hastings (HASTINGS, 1970); e (ii) a amostragem de *Gibbs*, que é essencialmente um caso especial da amostragem *Metropolis-Hastings*, uma vez que a probabilidade de aceitação é igual a um, e é atualmente o algoritmo de amostragem MCMC mais popular na literatura de inferência de confiabilidade Bayesiana (GEMAN; GEMAN, 1984; TANNER; WONG, 1987).

Nos últimos anos, uma plataforma de software foi implementada com vários programas independentes que incluem: BEAUti (*Bayesian Evolutionary Analysis Utility*), que é uma interface gráfica de usuário para configurar as análises (DRUMMOND *et al.*, 2012) e o BEAST (*Bayesian Evolutionary Analysis by Sampling Trees*) que executa análises MCMC e ferramentas de processamento como “LogAnalyser”, “LogCombiner”, “TreeAnnotator” (BOUCKAERT; HELED, 2014). Este software também executa outros pacotes de administrador para simular modelos estocásticos de dinâmica populacional, para inferir modelos de estruturas coalescentes, saltos reversíveis entre modelos de substituição, modelos com múltiplas espécies coalescentes

utilizando dados de SNP (do inglês, *Single Nucleotide Polymorphism*), entre outros (BOUCKAERT; ALVARADO-MORA; REBELLO PINHO, 2013; BRYANT *et al.*, 2012; STADLER *et al.*, 2013; VAUGHAN *et al.*, 2014; VAUGHAN; DRUMMOND, 2013; WU; SUCHARD; DRUMMOND, 2013).

7. Relógio molecular

Por meio do relógio molecular podemos estimar escalas de tempo evolucionárias a partir de dados genéticos e bioquímicos (BROMHAM *et al.*, 2018; HO *et al.*, 2016; LUO; HO, 2018). O uso de ferramentas estatísticas permite a análise de sequências genômicas e cálculo dos sinais de relação filogenética (YANG; RANNALA, 2012), das taxas de mutação e escalas de tempo evolutivas (BROMHAM; PENNY, 2003). O relógio molecular é uma ferramenta poderosa que descreve como os genomas de diferentes organismos divergem ao longo do tempo (LUO; HO, 2018), ou seja, estima-se uma relação proporcional entre a distância genética de duas espécies e o tempo decorrido desde a última vez que elas compartilharam um ancestral comum (MORGAN, 1998; BROMHAM *et al.*, 2018). O termo “relógio evolutivo molecular” foi proposto na década de 1960 por Zuckerkandl e Pauling (MORGAN, 1998; ZUCKERKANDL; PAULING, 1965), em que observaram que o grau de divergência das proteínas da hemoglobina entre espécies tinha uma relação linear com o tempo. Anos mais tarde, descobriu-se que as sequências de aminoácidos de várias proteínas diferentes mudavam a uma taxa linear ao longo do tempo, permitindo que os tempos de divergência fossem estimados para várias espécies (DICKERSON, 1971). Depois dos anos 70, o relógio molecular foi usado para estimar a taxa de evolução e a escala de tempo de vários patógenos com base em dados de sequências de DNA, RNA e sequências de proteínas (KUMAR, 2005). Assumir que a divergência de sequências de nucleotídeos ou aminoácidos ocorre constantemente ao longo do tempo é a maneira mais simples de estimar-se a taxa evolutiva (KIMURA, 1987), portanto, um relógio molecular relativamente constante em sequências de proteínas foi um forte suporte para a teoria neutra da evolução molecular (KIMURA, 1969). As mutações adaptativas são determinadas por vários parâmetros evolutivos, como a força da vantagem seletiva das mutações, o tamanho da população efetiva e as taxas de mutação adaptativa (KIMURA, 1983). Por isso, muita atenção tem sido dada às taxas de evolução constantes, e portanto, foi proposta a hipótese de um relógio molecular por Kimura (KIMURA, 1968), onde indicou que mutações neutras poderiam explicar uma taxa evolutiva constante (MORGAN, 1998; ZUCKERKANDL; PAULING, 1965). Em outras palavras, a teoria neutra prevê que a taxa de mudança evolutiva é constante entre as espécies, quando a duração de suas gerações é idêntica (KIMURA, 1987). Essa teoria se baseia na premissa de que a

mudança genética é o resultado de erros durante a replicação do genoma, e, portanto, espécies com gerações mais curtas devem acumular mudanças genéticas ao longo do tempo (LUO; HO, 2018). No entanto, esse efeito varia de acordo com os tipos de mutações de nucleotídeos (MOORJANI *et al.*, 2016). As sequências de DNA que codificam proteínas são divididas em dois tipos de sítios, “sinônimos”, que são aqueles que codificam posições degeneradas na tabela de códons, onde uma mudança não leva a uma substituição de aminoácido; e os “não sinônimos”, em que qualquer mudança levaria a substituições de aminoácidos (LI; WU; LOU, 1985). Conforme as técnicas de sequenciamento melhoram, as sequências de DNA que incluem regiões não codificantes ou apenas as regiões extraídas de múltiplos alinhamentos de genomas parciais ou completos são usadas para testar relógios moleculares neutros (ELANGO; THOMAS; YI, 2006; HUTTLEY; WAKEFIELD; EASTEAL, 2007).

Ao longo dos anos, várias formas de variação da taxa evolutiva foram reconhecidas (GAUT, 2011), devido ao fato de que diferentes genes e proteínas evoluem em taxas diferentes (DICKERSON, 1971). Esses efeitos genéticos podem ser causados por restrição seletiva entre genes e proteínas (KIMURA; OTA, 1974). Os “efeitos de sítio” podem ocorrer dentro de genes e proteínas, como por exemplo, alguns sítios de aminoácidos com importância funcional podem ser invariantes (LUO; HO, 2018). No entanto, a distribuição gamma é comumente usada para uma distribuição de probabilidade escolhida, porque pode assumir uma variedade de formas que representam diferentes graus de taxa de heterogeneidade entre sítios (YANG, 1996). Um desafio para o relógio molecular é a variação das taxas entre as linhagens. A heterogeneidade das taxas é conhecida como o “efeito linhagem” que pode ser promovido por diferenças a nível de espécie pelo mecanismo de reparo do DNA, da duração das gerações, entre outros (BROMHAM, 2009, 2011). A taxa evolutiva também pode ser correlacionada com o tamanho do genoma e escala de tempo (TOMOKO, 1995). É sabido que genomas de vírus de RNA evoluem a taxas muito elevadas em comparação com genomas nucleares de vertebrados (SANJUÁN *et al.*, 2010).

A aplicação prática de relógios moleculares teve suas complicações devido às variações nos cálculos das taxas evolutivas, e, portanto, foram desenvolvidos modelos realistas que levam em consideração essas variações (HO; DUCHÊNE, 2014; KUMAR; BLAIR HEDGES, 2016), onde é permitido a variação da taxa entre os diferentes ramos da filogenia (Figura 9). Vários modelos de relógio molecular foram desenvolvidos, e entre os mais conhecidos temos: o modelo estrito, modelo local, modelo relaxado autocorrelacionado e o modelo relaxado não correlacionado.

Modelo de relógio estrito: é o modelo mais simples. Assume-se que a taxa é constante ao longo de toda a filogenia (Figura 9a) e com um único parâmetro representado pela taxa de

evolução, que é expressada em substituições de nucleotídeos por sítio por ano. Este modelo assume que cada ramo da árvore filogenética evolui com a mesma taxa evolutiva, e, portanto, é um modelo de um parâmetro (taxa de conversão entre comprimentos de ramo e tempo de evolução) (FERREIRA; SUCHARD, 2008).

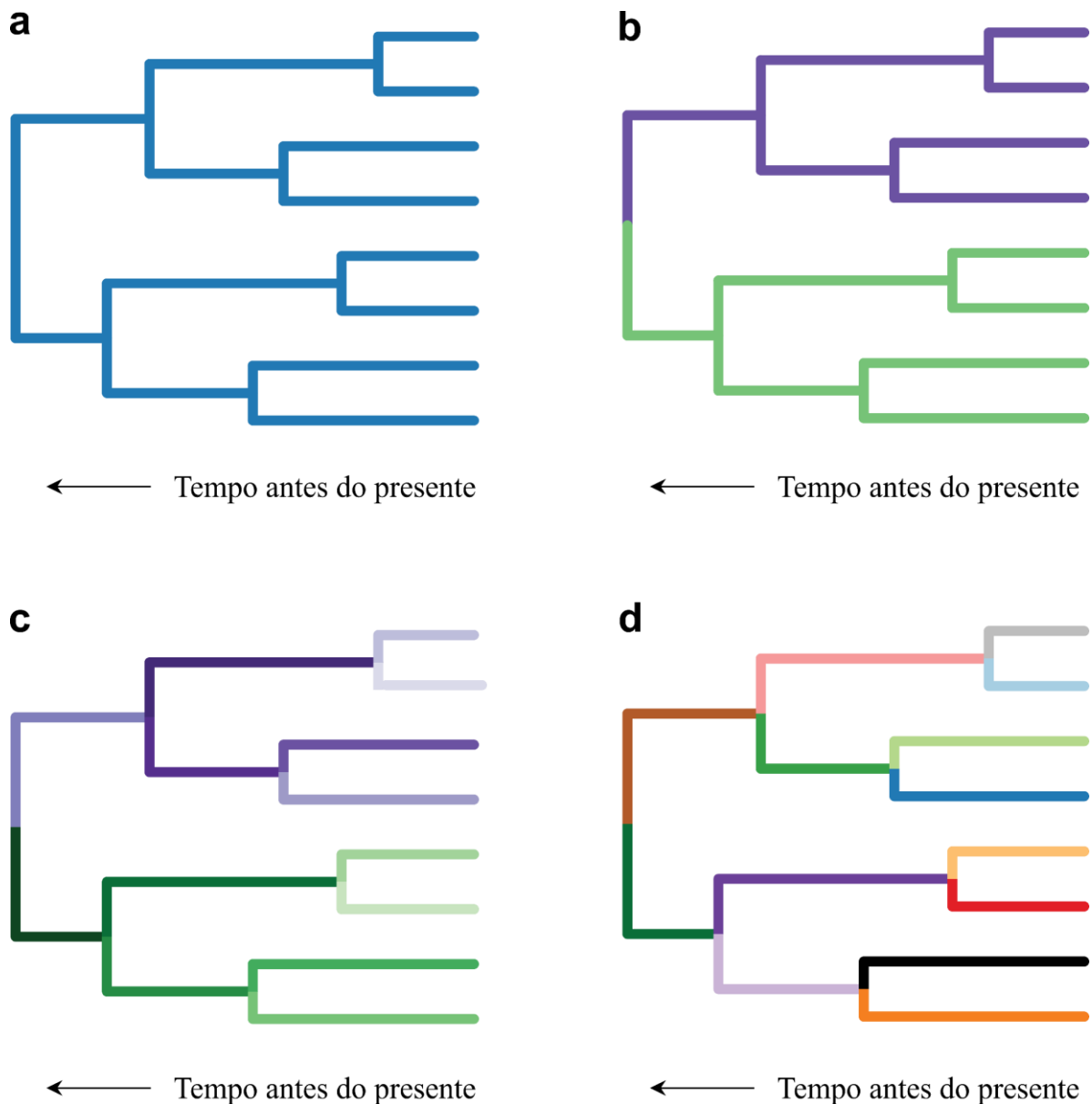
Modelo de relógio local: permite que ramos vizinhos compartilhem a mesma taxa evolutiva (YODER; YANG, 2000) (Figura 9b). Isso significa que clados/linhagens podem evoluir de acordo com diferentes taxas evolutivas, enquanto a constância da taxa é assumida no resto de árvore (YODER; YANG, 2000).

Modelo de relógio relaxado autocorrelacionado: é aquele que permite que a taxa evolua gradativamente ao longo do ramo (Figura 9c) (LEPAGE *et al.*, 2007; SANDERSON, 2002; THORNE; KISHINO; PAINTER, 1998).

Modelo de relógio relaxado não-correlacionado: supõe-se que as velocidades ao longo de vários ramos são desenhadas independentemente de uma distribuição de probabilidade escolhida (Figura 9d) (DRUMMOND *et al.*, 2006; RANNALA; YANG, 2007), ou seja, cada ramo de uma árvore filogenética tem sua própria taxa de evolução (DRUMMOND *et al.*, 2006). Portanto, estes modelos são chamados de não-correlacionados porque a taxa de evolução em um ramo não depende da taxa de qualquer um dos ramos vizinhos (DRUMMOND *et al.*, 2006), e a taxa de evolução ao longo do ramo pode ir de rápida para lenta ou vice-versa abruptamente. As distribuições admitidas no programa BEAUti são do tipo log-normal, exponencial e gama (DRUMMOND *et al.*, 2012).

Atualmente, estes dois últimos modelos de relógio são os mais utilizados pois permitem taxas evolutivas ao longo de cada ramo da árvore filogenética, são ricos em parâmetros e pode-se restringir a variação das taxas entre as linhagens.

Figura 9 - Modelos de relógios molecular.



(a) Modelo estrito, assume uma taxa constante de evolução em toda a árvore. (b) Modelo local, há uma mudança na taxa de evolução no TMRCA de cada conjunto de táxon definido (taxas de evolução separadas são assumidas para as metades superior e inferior da árvore). (c) Modelo relaxado autocorrelacionado, permitindo que as taxas de evolução mudam gradualmente ao longo dos ramos da árvore filogenética. (d) Modelo relaxado não correlacionado, suponha que as taxas específicas do ramo sejam traçadas independentemente de uma distribuição subjacente. Cores e tons diferentes indicam taxas diferentes, com cores semelhantes indicando taxas semelhantes. Adaptado de LUO e HO (2018) e <https://beast.community/clocks>.

Por meio de métodos estatísticos é possível comparar o melhor ajuste entre os diferentes modelos de relógio (BAELE *et al.*, 2013; DOS REIS *et al.*, 2018) e identificar se o modelo escolhido descreve adequadamente a heterogeneidade de frequências presentes em um conjunto de dados (DUCHÊNE *et al.*, 2015). Na prática, o uso de modelos de relógio relaxado tornou-se rotina, pois o relógio estrito desempenha principalmente um papel como uma hipótese nula ao testar-se a variação nas taxas evolutivas (LUO; HO, 2018). Por fim, foram desenvolvidas

abordagens para selecionar modelos filogenéticos, tais como amostragem de caminhos (*path sampling*, PS) (LARTILLOT; PHILIPPE, 2006), amostragem por etapas (*stepping-stone sampling*, SS) (XIE *et al.*, 2011) e amostragem generalizadas por etapas (FAN *et al.*, 2011). As abordagens PS e SS têm se mostrado bons estimadores do modelo demográfico e do relógio molecular.

Por tanto, os objetivos deste trabalho foram: (i) estudar molecularmente os vírus que infectam importantes dioscoreáceas presentes na região Nordeste do Brasil. Para tanto, foram coletadas amostras foliares de seis espécies de *Dioscorea* e submetidas ao sequenciamento de alto desempenho pela plataforma Illumina; (ii) estimar a prevalência de espécies de vírus de RNA em *Dioscorea cayennensis-rotundata*, (iii) estudar sequências genômicas completas de YMV utilizando dados de HTS e Sanger oriundas de RNA total extraído da espécie *Dioscorea cayennensis-rotundata*; (iv) realizar estudos filodinâmicos bayesianos usando sequências do gene CP para calcular o tempo mais recente do último ancestral comum (TMRCA) e conhecer a possível origem geográfica de YMV. Por fim, buscou-se (v) relatar e caracterizar a sequência genômica de YMMV isolada de plantas de *D. alata* a partir de dados de HTS e com base na sequência de nt do gene CP de YMMV propor a reestruturação dos grupos filogenéticos.

8. Referências Bibliográficas

AAS, P. A. *et al.* Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. **Nature**, v. 421, n. 6925, p. 859–863, 2003. DOI: 10.1038/nature01363.

ADACHI, S. *et al.* Effects of infection by Turnip mosaic virus on the population growth of generalist and specialist aphid vectors on turnip plants. **PLoS ONE**, v. 13, n. 7, p. 1–8, 2018. DOI: 10.1371/journal.pone.0200784.

ADAMS, I.; FOX, A. Diagnosis of Plant Viruses Using Next- Generation Sequencing and Metagenomic Analysis. **Current Research Topics in Plant Virology**, p. 1–335, 2016. DOI: 10.1007/978-3-319-32919-2_14.

ADAMS, I. P. *et al.* Next-generation sequencing and metagenomic analysis: A universal diagnostic tool in plant virology. **Molecular Plant Pathology**, v. 10, n. 4, p. 537–545, 2009. DOI: 10.1111/j.1364-3703.2009.00545.x.

ADAMS, M. J. *et al.* **Virus and Virus-like Disease of Major Crops in Developing Countries**. Volume I. Ed. Israel and India: G. Loebenstein and G. Thottappilly, 2003.

ADAMS, M. J.; SWABY, A. G.; JONES, P. Confirmation of the transmission of barley yellow mosaic virus (BaYMV) by the fungus *Polymyxa graminis*. **Annals of Applied Biology**, v. 112, n. 1, p. 133–141, 1988. DOI: 10.1111/j.1744-7348.1988.tb02048.x.

ADAMS, M. J.; ANTONIW, J. F.; BEAUDOIN, F. Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. **Molecular Plant Pathology**, v. 6, n. 4, p. 471–487, 2005. DOI: 10.1111/j.1364-3703.2005.00296.x.

ADENIJI, M. O. *et al.* Yield loss in guinea yam (*dioscorea rotundata* poir.) due to infection by yam mosaic virus (YMV) genus potyvirus. **Ife Journal of Science**, v. 14, n. 2, p. 237–244, 2012.

AFIAHAYATI; SATO, K; SAKAKIBARA, Y. MetaVelvet-SL: An extension of the Velvet assembler to a de novo metagenomic assembler utilizing supervised learning. **DNA Research**, v. 22, n. 1, p. 69–77, 2015. DOI: 10.1093/dnares/dsu041.

AGUDELO-ROMERO, P. *et al.* Changes in the gene expression profile of *Arabidopsis thaliana* after infection with Tobacco etch virus. **Virology Journal**, v. 5, p. 1–11, 2008. DOI: 10.1186/1743-422X-5-92.

AL RWAHNIH, M. *et al.* Complete Genome Sequence of a Novel Vitivirus Isolated from Grapevine. **Journal of Virology**, v. 86, n. 17, p. 9545–9545, 2012. DOI: 10.1128/jvi.01444-12.

AL RWAHNIH, M. *et al.* Comparison of next-generation sequencing versus biological indexing for the optimal detection of viral pathogens in grapevine. **Phytopathology**, v. 105, n. 6, p. 758–763, 2015. DOI: 10.1094/PHYTO-06-14-0165-R.

ALA-POIKELA, M. *et al.* T. Helper Component Proteinase of the Genus Potyvirus Is an Interaction Partner of Translation Initiation Factors eIF(iso)4E and eIF4E and Contains a 4E Binding Motif. **Journal of Virology**, v. 85, n. 13, p. 6784–6794, 2011. DOI: 10.1128/jvi.00485-11.

ALEMAN-VERDAGUER, M. E. *et al.* Analysis of the sequence diversity of the P1, HC, P3, NIb and CP genomic regions of several yam mosaic potyvirus isolates: Implications for the intraspecies molecular diversity of potyviruses. **Journal of General Virology**, v. 78, n. 6, p. 1253–1264, 1997. DOI: 10.1099/0022-1317-78-6-1253.

ALEMAN, M. E. *et al.* The complete nucleotide sequence of yam mosaic virus (Ivory Coast isolate) genomic RNA. **Archives of Virology**, v. 141, n. 7, p. 1259–1278, 1996. DOI: 10.1007/BF01718829.

AMUSA, N. A. *et al.* Yam diseases and its management in Nigeria. **African Journal of Biotechnology**, v. 2, n. 12, p. 514–525, 2003. DOI: 10.5897/ajb2003.000-1099.

ANANDALAKSHMI, R. *et al.* A viral suppressor of gene silencing in plants. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 22, p. 13079–13084, 1998. DOI: 10.1073/pnas.95.22.13079.

ANDRES, C.; ADEOLUWA, O. O.; BHULLAR, G. S. Yam (*Dioscorea* spp.). **Encyclopedia of Applied Plant Sciences**, v. 3, p. 435–441, 2016. DOI: 10.1016/B978-0-12-394807-6.00177-5.

ANDRADE, G.P. *et al.* Fitopatol.bras. 32 (supl): 322. 2007.

ANZALA, F. *et al.* Molecular characterization of Secoviridae and Potexvirus infecting yams (*Dioscorea* spp). [S. l.], p. 3, [s.d.].

ARANDA, M.; MAULE, A. Virus-induced host gene shutoff in animals and plants. **Virology**, v. 243, n. 2, p. 261–267, 1998. DOI: 10.1006/viro.1998.9032.

ARAVIND, L., KOONIN, E. V. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. **Genome Biol**, v. 2, n. 3, p. research0007.1–research0007.8., 2001. DOI: 10.1186/gb-2001-2-3-research0007.

ARENAS, M. *et al.* Mutation and recombination in pathogen evolution: Relevance, methods and controversies. **Infection, Genetics and Evolution**, v. 63, p. 295–306, 2018. DOI: 10.1016/j.meegid.2017.09.029.

ARIAS, M. C. *et al.* Sunflower chlorotic mottle virus in compatible interactions with sunflower: ROS generation and antioxidant response. **European Journal of Plant Pathology**, v. 113, n. 3, p. 223–232, 2005. DOI: 10.1007/s10658-005-7559-5.

ASALA, S. W.; ALEGBEJO, M. D. Effects of serial planting of seed yam tubers on virus incidence and yam tuber degeneration. **African Crop Science Journal**, v. 24, n. 4, p. 341, 2016. DOI: 10.4314/acsj.v24i4.1.

ASALA, S. *et al.* Distribution and incidence of viruses infecting yam (*Dioscorea* spp.) in Nigeria. **Global journal of bioscience and biotechnology**, v. 1, n. 2, p. 163–167, 2012.

ATREYA, C. D. *et al.* Site-directed mutations in the potyvirus HC-PRO gene affect helper component activity, virus accumulation, and symptom expression in infected tobacco plants. **Virology**, v. 191, n. 1, p. 106–111, 1992. DOI: 10.1016/0042-6822(92)90171-K.

ATREYA, C. D.; PIRONE, T. P. Mutational analysis of the helper component-proteinase gene of a potyvirus: Effects of amino acid substitutions, deletions, and gene replacement on virulence and aphid transmissibility. **Proceedings of the National Academy of Sciences of the United States of America**, v. 90, n. 24, p. 11919–11923, 1993. DOI: 10.1073/pnas.90.24.11919.

AZAD, M. A. K.; AMIN, L.; SIDIK, N. M. Gene technology for papaya ringspot virus disease management. **The Scientific World Journal**, v. 2014, 2014. DOI: 10.1155/2014/768038.

BAELE, G. *et al.* Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. **Molecular Biology and Evolution**, v. 30, n. 2, p. 239–243, 2013. DOI: 10.1093/molbev/mss243.

BAK, A. *et al.* Ethylene signaling mediates potyvirus spread by aphid vectors. **Oecologia**, v. 190, n. 1, p. 139–148, 2019. DOI: 10.1007/s00442-019-04405-0.

BALLUT, L. *et al.* HcPro, a multifunctional protein encoded by a plant RNA virus, targets the 20S proteasome and affects its enzymic activities. **Journal of General Virology**, v. 86, n. 9, p. 2595–2603, 2005. DOI: 10.1099/vir.0.81107-0.

BANKEVICH, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. **Journal of Computational Biology**, v. 19, n. 5, p. 455–477, 2012. DOI: 10.1089/cmb.2012.0021.

BARBA, M.; CZOSNEK, H.; HADIDI, A. Historical perspective, development and applications of next-generation sequencing in plant virology. **Viruses**, v. 6, n. 1, p. 106–136, 2013. DOI: 10.3390/v6010106.

BARRERO, R. A. *et al.* An internet-based bioinformatics toolkit for plant biosecurity diagnosis and surveillance of viruses and viroids. **BMC Bioinformatics**, v. 18, n. 1, p. 1–12, 2017. DOI: 10.1186/s12859-016-1428-4.

BEHBOODIAN, B. *et al.* Postharvest analysis of lowland transgenic tomato fruits harboring hpRNAi- ACO1 construct. **The Scientific World Journal**, v. 2012, n. 2002, 2012. DOI: 10.1100/2012/439870.

BELSHAM, G.J.; SONENBERG, N. Picornavirus RNA translation: Roles for cellular proteins. **Trends in Microbiology**, v. 8, n. 7, p. 330–335, 2000. DOI: 10.1016/S0966-842X(00)01788-1.

BENTLEY, D. R. *et al.* Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry. **Nature**, v. 456, n. 7218, p. 53–59, 2008. DOI: 10.1038/nature07517.

BLANC, S. *et al.* Mutations in the potyvirus helper component protein: Effects on interactions with virions and aphid stylets. **Journal of General Virology**, v. 79, n. 12, p. 3119–3122, 1998. DOI: 10.1099/0022-1317-79-12-3119.

BLANC, S. *et al.* A specific interaction between coat protein and helper component correlates with aphid transmission of a potyvirus. **Virology**, v. 231, n. 1, p. 141–147, 1997. DOI: 10.1006/viro.1997.8521.

BLAWID, R.; SILVA, J. M. F.; NAGATA, T. Discovering and sequencing new plant viral genomes by next-generation sequencing: description of a practical pipeline. **Annals of Applied Biology**, v. 170, n. 3, p. 301–314, 2017. DOI: 10.1111/aab.12345.

BLUA, M. J.; PERRING, T. M.; MADORE, M. A. Plant virus-induced changes in aphid population development and temporal fluctuations in plant nutrients. **Journal of Chemical Ecology**, v. 20, n. 3, p. 691–707, 1994. DOI: 10.1007/BF02059607.

BOETZER, M.; PIROVANO, W. Toward almost closed genomes with GapFiller. **Genome Biology**, v. 13, n. 6, p. R56, 2012. DOI: 10.1186/gb-2012-13-6-r56.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: A flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 2014. DOI: 10.1093/bioinformatics/btu170.

BÖMER, M. *et al.* Tissue culture and next-generation sequencing: A combined approach for detecting yam (*Dioscorea* spp.) viruses. **Physiological and Molecular Plant Pathology**, v. 105, p. 54–66, 2019. DOI: 10.1016/j.pmpp.2018.06.003.

BÖMER, M. *et al.* A sequence-independent strategy for amplification and characterisation of episomal badnavirus sequences reveals three previously uncharacterised yam badnaviruses. **Viruses**, v. 8, n. 7, p. 14–16, 2016. DOI: 10.3390/v8070188.

BOQUEL, S.; GIORDANENGO, P.; AMELINE, Arnaud. Divergent effects of PVY-infected potato plant on aphids. **European Journal of Plant Pathology**, v. 129, n. 4, p. 507–510, 2011. DOI: 10.1007/s10658-010-9732-8.

BOS, L.; DIAZ-RUIZ, J. R.; MAAT, D. Z. Further characterization of celery latent virus. **Netherlands Journal of Plant Pathology**, v. 84, n. 2, p. 61–79, 1978. DOI: 10.1007/BF01976409.

BOUCKAERT, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. **PLoS Computational Biology**, v. 15, n. 4, p. 1–28, 2019. DOI: 10.1371/journal.pcbi.1006650.

BOUCKAERT, R.; ALVARADO-MORA, M. V.; REBELLO PINHO, J. R. Evolutionary rates and HBV: Issues of rate estimation with Bayesian molecular methods. **Antiviral Therapy**, v. 18, n. 3 PARTB, p. 497–503, 2013. DOI: 10.3851/IMP2656.

BOUCKAERT, R.; HELED, J. DensiTree 2: Seeing Trees Through the Forest. **bioRxiv**, p. 012401, 2014. DOI: 10.1101/012401.

BOUSALEM, M.; DOUZERY, E. J. P.; FARGETTE, D. High genetic diversity, distant phylogenetic relationships and intraspecies recombination events among natural populations of Yam mosaic virus: A contribution to understanding potyvirus evolution. **Journal of General Virology**, v. 81, n. 1, p. 243–255, 2000. DOI: 10.1099/0022-1317-81-1-243.

BOUSALEM, M. *et al.* Origin, world-wide dispersion, bio-geographical diversification, radiation and recombination: An evolutionary history of Yam mild mosaic virus (YMMV).

Infection, Genetics and Evolution, v. 3, n. 3, p. 189–206, 2003. DOI: 10.1016/S1567-1348(03)00085-6.

BOUSALEM, M. *et al.* Dilemmas caused by endogenous pararetroviruses regarding the taxonomy and diagnosis of yam (*Dioscorea* spp.) badnaviruses: Analyses to support safe germplasm movement. **Archives of Virology**, v. 154, n. 2, p. 297–314, 2009. DOI: 10.1007/s00705-009-0311-2.

BRAGARD, C. *et al.* Status and prospects of plant virus control through interference with vector transmission. **Annual Review of Phytopathology**, v. 51, p. 177–201, 2013. DOI: 10.1146/annurev-phyto-082712-102346.

BRANDES, J.; LUISONI, E. Untersuchungen über einige Eigenschaften von zwei gestreckten Sellerieviren. **Journal of Phytopathology**, v. 57, n. 3, p. 277–288, 1966. DOI: <https://doi.org/10.1111/j.1439-0434.1966.tb02282.x>.

BRIDDON, R. *et al.* Analysis of the sequence of *Dioscorea alata* bacilliform virus; comparison to other members of the badnavirus group. **Virus Genes**, v. 18, n. 3, p. 277–283, 1999. DOI: 10.1023/A:1008076420783.

BRISON, M. *et al.* Effect of cryopreservation on the sanitary state of a cv *Prunus* rootstock experimentally contaminated with Plum Pox Potyvirus. **Plant Science**, v. 123, n. 1–2, p. 189–196, 1997. DOI: 10.1016/S0168-9452(97)04581-0.

BROMHAM, L. Why do species vary in their rate of molecular evolution? **Biology Letters**, v. 5, n. 3, p. 401–404, 2009. DOI: 10.1098/rsbl.2009.0136.

BROMHAM, L. The genome as a life-history character: Why rate of molecular evolution varies between mammal species. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 366, n. 1577, p. 2503–2513, 2011. DOI: 10.1098/rstb.2011.0014.

BROMHAM, L. *et al.* Bayesian molecular dating: opening up the black box. **Biological Reviews**, v. 93, n. 2, p. 1165–1191, 2018. DOI: 10.1111/brv.12390.

BROMHAM, L.; PENNY, D. The modern molecular clock. **Nature Reviews Genetics**, v. 4, n. 3, p. 216–224, 2003. DOI: 10.1038/nrg1020.

BRYANT, D. *et al.* Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. **Molecular Biology and Evolution**, v. 29, n. 8, p. 1917–1932, 2012. DOI: 10.1093/molbev/mss086.

CARD, S. D.; PEARSON, M. N.; CLOVER, G. R. G. Plant pathogens transmitted by pollen. **Australasian Plant Pathology**, v. 36, n. 5, p. 455–461, 2007. DOI: 10.1071/AP07050.

CARRINGTON, J. C.; DOUGHERTY, W. G. Small Nuclear Inclusion Protein Encoded by a Plant Potyvirus Genome Is a Protease. **Journal of Virology**, v. 61, n. 8, p. 2540–2548, 1987. DOI: 10.1128/jvi.61.8.2540-2548.1987.

CASTEEL, C. L. *et al.* The NIa-Pro protein of Turnip mosaic virus improves growth and reproduction of the aphid vector, *Myzus persicae* (green peach aphid). **Plant Journal**, v. 77, n. 4, p. 653–663, 2014. DOI: 10.1111/tpj.12417.

CASTLE, S. J.; BERGER, P. H. Rates of growth and increase of *Myzus persicae* on virus-infected potatoes according to type of virus-vector relationship. **Entomologia Experimentalis et Applicata**, v. 69, n. 1, p. 51–60, 1993. DOI: 10.1111/j.1570-7458.1993.tb01727.x.

CHAI, M. *et al.* P3N-PIPO Interacts with P3 via the Shared N-Terminal Domain To Recruit Viral Replication Vesicles for Cell-to-Cell Movement. **Journal of Virology**, v. 94, n. 8, 2020. DOI: 10.1128/jvi.01898-19.

CHAN, Y. *et al.* A diet containing yam reduces the cognitive deterioration and brain lipid peroxidation in mice with senescence accelerated. **International Journal of Food Science and Technology**, v. 39, n. 1, p. 99–107, 2004. DOI: 10.1046/j.0950-5423.2003.00751.x.

CHANDRASEKARA, A.; JOSHEPH KUMAR, T. Roots and tuber crops as functional foods: A review on phytochemical constituents and their potential health benefits. **International Journal of Food Science**, v. 2016, 2016. DOI: 10.1155/2016/3631647.

CHAPMAN, E. J. *et al.* Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. **Genes and Development**, v. 18, n. 10, p. 1179–1186, 2004. DOI: 10.1101/gad.1201204.

CHELLEMI, D. O. *et al.* Widespread occurrence and low genetic diversity of Colombian datura virus in *Brugmansia* suggest an anthropogenic role in virus selection and spread. **Plant Disease**, v. 95, n. 6, p. 755–761, 2011. DOI: 10.1094/PDIS-09-10-0654.

CHEN, H. L. *et al.* Effects of Taiwanese yam (*Dioscorea japonica* Thunb var. *pseudojaponica* Yamamoto) on upper gut function and lipid metabolism in Balb/c mice. **Nutrition**, v. 19, n. 7–8, p. 646–651, 2003. DOI: 10.1016/S0899-9007(03)00058-3.

CHEN, J. *et al.* Characterisation of a potyvirus and a potexvirus from Chinese scallion. **Archives of Virology**, v. 147, n. 4, p. 683–693, 2002. DOI: 10.1007/s007050200018.

CHEN, J. *et al.* Molecular analysis of barley yellow mosaic virus isolates from China. **Virus Research**, v. 64, n. 1, p. 13–21, 1999. DOI: 10.1016/S0168-1702(99)00076-3.

CHEN, Y. K. *et al.* High frequency induction of RNA-mediated resistance against Cucumber mosaic virus using inverted repeat constructs. **Molecular Breeding**, v. 14, n. 3, p. 215–226, 2004. DOI: 10.1023/B:MOLB.0000047769.82881.f5.

CHENG, X.; WANG, A. The Potyvirus Silencing Suppressor Protein VPg Mediates Degradation of. **Journal of Virology**, v. 91, n. 1, p. e01478-16, 2017.

CHOI, D. *et al.* Two novel poty-like viruses identified from the transcriptome data of purple witchweed (*Striga hermonthica*). **Acta virologica**, v. 65, n. 04, p. 365–372, 2022. DOI: 10.4149/av_2021_402.

CHOUDHARY, S. K. *et al.* Comprehensive evaluation of anti-hyperglycemic activity of fractionated momordica charantia seed extract in alloxan-induced diabetic rats. **Evidence-based Complementary and Alternative Medicine**, v. 2012, p. 10, 2012. DOI: 10.1155/2012/293650.

CHUNG, B.Y. W. *et al.* An overlapping essential gene in the Potyviridae. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 15, p. 5897–5902, 2008. DOI: 10.1073/pnas.0800468105.

CLARK, C. *et al.* Sweetpotato viruses: 15 years of progress on understanding and managing complex diseases. **Plant Disease**, v. 96, n. 2, p. 168–185., 2012. DOI: <http://dx.doi.org/10.1094/PDIS-07-11-0550>.

COMPEAU, P. E. C.; PEVZNER, P. A.; TESLER, G. How to apply de Bruijn graphs to genome assembly. **Nature Biotechnology**, v. 29, n. 11, p. 987–991, 2011. DOI: 10.1038/nbt.2023.

CONG, Q. Q. *et al.* Evaluation of Potato virus X mild mutants for cross protection against severe infection in China. **Virology Journal**, v. 16, n. 1, p. 1–13, 2019. DOI: 10.1186/s12985-019-1143-7.

COUTO, R. S. *et al.* Time calibrated tree of Dioscorea (Dioscoreaceae) indicates four origins of yams in the Neotropics since the Eocene. **Botanical Journal of the Linnean Society**, v. 188, n. 2, p. 144–160, 2018. DOI: 10.1093/botlinnean/boy052.

COUTO, R. S.; FRAGA, F.R.M. 2020. Dioscoreaceae in **Flora do Brasil 2020**. Jardim Botânico do Rio de Janeiro. Disponível em: <<http://reflora.jbrj.gov.br/reflora/floradobrasil/FB104>>. Acesso em: 14 Jan. 2022.

COSTA, D.P. *et al.* XVII Enc. Nac. Metodologia e Gestão de Lab. Da Embrapa. 2013.

CUI, H.; WANG, A. Plum Pox Virus 6K1 Protein Is Required for Viral Replication and Targets the Viral Replication Complex at the Early Stage of Infection. **Journal of Virology**, v. 90, n. 10, p. 5119–5131, 2016. DOI: 10.1128/jvi.00024-16.

CUI, X. *et al.* The C-terminal region of the Turnip mosaic virus P3 protein is essential for viral infection via targeting P3 to the viral replication complex. **Virology**, v. 510, n. July, p. 147–155, 2017. DOI: 10.1016/j.virol.2017.07.016.

DESBIEZ, C. *et al.* Increase in Zucchini yellow mosaic virus Symptom Severity in Tolerant Zucchini Cultivars Is Related to a Point Mutation in P3 Protein and Is Associated with a Loss of Relative Fitness on Susceptible Plants. **Phytopathology**, v. 93, n. 12, p. 1478–1484, 2003. DOI: 10.1094/PHYTO.2003.93.12.1478.

DEY, K.K. *et al.* Air potato (*Dioscorea bulbifera*) plants displaying virus-like symptoms are co-infected with a novel potyvirus and a novel ampelovirus. **Virus Genes**, v. 55, n. 1, p. 117–121, 2019. DOI: 10.1007/s11262-018-1616-6.

DIAZ-LARA, A. *et al.* Two novel negative-sense RNA viruses infecting grapevine are members of a newly proposed genus within the family phenuiviridae. **Viruses**, v. 11, n. 8, 2019. DOI: 10.3390/v11080685.

DÍAZ-VIVANCOS, P. *et al.* Alteration in the chloroplastic metabolism leads to ROS accumulation in pea plants in response to plum pox virus. **Journal of Experimental Botany**, v. 59, n. 8, p. 2147–2160, 2008. DOI: 10.1093/jxb/ern082.

DICKERSON, R. E. The structure of cytochrome c and the rates of molecular evolution. **Journal of Molecular Evolution**, v. 1, n. 1, p. 26–45, 1971. DOI: 10.1007/BF01659392.

DIETZGEN, R. G.; MANN, K. S.; JOHNSON, K. N. Plant virus-insect vector interactions: Current and potential future research directions. **Viruses**, v. 8, n. 11, p. 1–21, 2016. DOI: 10.3390/v8110303.

DOGIMONT, C. *et al.* The Vat locus encodes for a CC-NBS-LRR protein that confers resistance to *Aphis gossypii* infestation and *A. gossypii*-mediated virus resistance. **Plant Journal**, v. 80, n. 6, p. 993–1004, 2014. DOI: 10.1111/tpj.12690.

DOMBROVSKY, A. *et al.* In vitro association between the helper component-proteinase of zucchini yellow mosaic virus and cuticle proteins of *Myzus persicae*. **Journal of General Virology**, v. 88, n. 5, p. 1602–1610, 2007. DOI: 10.1099/vir.0.82769-0.

DOMBROVSKY, A.; REINGOLD, V.; ANTIGNUS, Y. Ipomovirus - an atypical genus in the family Potyviridae transmitted by whiteflies. **Pest Management Science**, v. 70, n. 10, p. 1553–1567, 2014. DOI: 10.1002/ps.3735.

DONAIRE, L. *et al.* Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. **Virology**, v. 392, n. 2, p. 203–214, 2009. DOI: 10.1016/j.virol.2009.07.005.

DOS REIS, M. *et al.* Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: Primates as a test case. **Systematic Biology**, v. 67, n. 4, p. 594–615, 2018. DOI: 10.1093/sysbio/syy001.

DREHER, T. W.; MILLER, W. A. Translational control in positive strand RNA plant viruses. **Virology**, v. 344, n. 1, p. 185–197, 2006. DOI: 10.1016/j.virol.2005.09.031.

DRUMMOND, A. J. *et al.* Relaxed phylogenetics and dating with confidence. **PLoS Biology**, v. 4, n. 5, p. 699–710, 2006. DOI: 10.1371/journal.pbio.0040088.

DRUMMOND, A. J. *et al.* Bayesian phylogenetics with BEAUti and the BEAST 1.7. **Molecular Biology and Evolution**, v. 29, n. 8, p. 1969–1973, 2012. DOI: 10.1093/molbev/mss075.

DUCHÊNE, D. A. *et al.* Evaluating the adequacy of molecular clock models using posterior predictive simulations. **Molecular Biology and Evolution**, v. 32, n. 11, p. 2986–2995, 2015. DOI: 10.1093/molbev/msv154.

EAGLES, R. M. *et al.* Characterization of NTPase, RNA-binding and RNA-helicase Activities of the Cytoplasmic Inclusion Protein of Tamarillo Mosaic Potyvirus. **European Journal of Biochemistry**, v. 224, n. 2, p. 677–684, 1994. DOI: 10.1111/j.1432-1033.1994.t01-1-00677.x.

ECKEL, R.V. W.; LAMPERT, E. P. Relative attractiveness of tobacco etch virus-infected and healthy flue-cured tobacco plants to aphids (Homoptera: Aphididae). **Journal of Economic Entomology**, v. 89, n. 4, p. 1017–1027, 1996. DOI: 10.1093/jee/89.4.1017.

EDWARDSON, J. R. Inclusion bodies. **Arch Virol**, v. 5, p. 25–30, 1992.

EDWARDSON, J. R.; PURCIFULL, D. E.; CHRISTIE, R. G. Structure of cytoplasmic inclusions in plants infected with rod-shaped viruses. **Virology**, v. 34, n. 2, p. 250–263, 1968. DOI: 10.1016/0042-6822(68)90234-1.

EDWARDSON, J. R. Cylindrical inclusions in the cytoplasm of leaf cells infected with tobacco etch virus. **Science**, v. 153, n. 3738, p. 883–884, 1966. DOI: 10.1126/science.153.3738.883.

EDWARSON, F. R.; CHRISTINE, R. G.; KO, N. J. Potyvirus Cylindrical Inclusion-Subdivision-IV. **Phytopathology**, v. 74, n. 9, p. 1111–1114, 1984. DOI: 10.1094/Phyto-74-1111.

EIGENBRODE, S. D. *et al.* Volatiles from potato plants infected with potato leafroll virus attract and arrest the virus vector, *Myzus persicae* (Homoptera: Aphididae). **Proceedings of the Royal Society B: Biological Sciences**, v. 269, n. 1490, p. 455–460, 2002. DOI: 10.1098/rspb.2001.1909.

ELANGO, N.; THOMAS, J. W.; YI, S. V. Variable molecular clocks in hominoids. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 5, p. 1370–1375, 2006. DOI: 10.1073/pnas.0510716103.

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. **Archives of Virology**, v. 153, n. 12, p. 2263–2272, 2008. a. DOI: 10.1007/s00705-008-0258-8.

ENI, A. O. *et al.* Characterization of cucumber mosaic virus isolated from yam (*Dioscorea* spp.) in West Africa. **African Journal of Biotechnology**, v. 12, n. 22, p. 3472–3480, 2013. DOI: 10.5897/AJB2013.12303.

ENI, A. O. *et al.* First report of Cucumber mosaic virus in yams (*Dioscorea* spp.) in Ghana, Togo, and Republic of Benin in West Africa. **Plant Disease**, v. 92, n. 5, p. 833, 2008. b. DOI: 10.1094/PDIS-92-5-0833B.

ESKELIN, K. *et al.* Potyviral VPg Enhances Viral RNA Translation and Inhibits Reporter mRNA Translation In Planta. **Journal of Virology**, v. 85, n. 17, p. 9210–9221, 2011. DOI: 10.1128/jvi.00052-11.

FAHIM, M. *et al.* RNA derived from viral NIa gene confers immunity to wheat streak mosaic virus infection in transgenic wheat plants. **Plant Biotechnology Journal**, v. 8, n. 7, p. 821–834, 2010. DOI: 10.1111/j.1467-7652.2010.00513.x.

FAN, Y. *et al.* Choosing among partition models in Bayesian phylogenetics. **Molecular Biology and Evolution**, v. 28, n. 1, p. 523–532, 2011. DOI: 10.1093/molbev/msq224.

FAOSTAT (Food and Agriculture Organization of the United Nations). **FAOSTAT Statistical Database**. Disponível em: <<http://www.fao.org/faostat/en/#data/QC>>. Acesso em: 20 jan. 2021.

FELLERS, J. P. *et al.* The complete genome sequence of Triticum mosaic virus, a new wheat-infecting virus of the High Plains. **Archives of virology**, v. 154, n. 9, p. 1511–1515, 2009. DOI: 10.1007/s00705-009-0462-1.

FELLERS, J. P. *et al.* The Potato virus Y MSNR NIb-replicase is the elicitor of a veinal necrosis-hypersensitive response in root knot nematode resistant tobacco. **Molecular Plant Pathology**, v. 3, n. 3, p. 145–152, 2002. DOI: <https://doi.org/10.1046/j.1364-3703.2002.00106.x>.

FELSENSTEIN, J. Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. **Journal of Molecular Evolution**, v. 17, p. 368–376, 1981. DOI: 10.1007/BF01734359.

FERERES, A.; MORENO, A. Behavioural aspects influencing plant virus transmission by homopteran insects. **Virus Research**, v. 141, n. 2, p. 158–168, 2009. DOI: 10.1016/j.virusres.2008.10.020.

FERNÁNDEZ-CALVINO, L. *et al.* The helper-component protease transmission factor of tobacco etch potyvirus binds specifically to an aphid ribosomal protein homologous to the laminin receptor precursor. **Journal of General Virology**, v. 91, n. 11, p. 2862–2873, 2010. DOI: 10.1099/vir.0.022335-0.

FERNÁNDEZ-CALVINO, L. *et al.* Comparison of Potato Virus Y and Plum Pox Virus transmission by two aphid species in relation to their probing behavior. **Phytoparasitica**, v. 34, n. 3, p. 315–324, 2006. DOI: 10.1007/BF02980959.

FERNÁNDEZ, A. *et al.* The motif V of plum pox potyvirus CI RNA helicase is involved in NTP hydrolysis and is essential for virus RNA replication. **Nucleic Acids Research**, v. 25, n. 22, p. 4474–4480, 1997. DOI: 10.1093/nar/25.22.4474.

FERREIRA, M. A. R.; SUCHARD, M. A. Bayesian analysis of elapsed times in continuous-time Markov chains. **Canadian Journal of Statistics**, v. 36, n. 3, p. 355–368, 2008. DOI: 10.1002/cjs.5550360302.

FERRO, D. N.; MACKENZIE, J. D.; MARGOLIES, D. C. Effect of Mineral Oil and a Systemic Insecticide on Field Spread of Aphid-Borne Maize Dwarf Mosaic Virus in Sweet Corn1. [*S. l.*], v. 73, n. 5, p. 730–735, 1980. DOI: <https://doi.org/10.1093/jee/73.5.730>.

FILLOUX, D. *et al.* Metagenomic discovery, worldwide distribution and genetic diversity of novel macluraviruses infecting yams (*Dioscorea* spp.). **15èmes Rencontres de Virologie Végétale**, Aussois –France, 2015.

FRENCH, R.; STENGER, D. C. Genome sequences of Agropyron mosaic virus and Hordeum mosaic virus support reciprocal monophyly of the genera Potyvirus and Rymovirus in the family Potyviridae. **Archives of Virology**, v. 150, n. 2, p. 299–312, 2005. DOI: 10.1007/s00705-004-0396-6.

FROISSART, R.; MICHALAKIS, Y.; BLANC, S. Helper component-transcomplementation in the vector transmission of plant viruses. **Phytopathology**, v. 92, n. 6, p. 576–579, 2002. DOI: 10.1094/PHYTO.2002.92.6.576.

FUENTES, S. *et al.* Potato virus Y; the Andean connection. **Virus Evolution**, v. 5, n. 2, 2019. DOI: 10.1093/ve/vez037.

FUJI, S.; NAKAMAE, H. Complete nucleotide sequence of the genomic RNA of a mild strain of Japanese yam mosaic potyvirus. **Archives of Virology**, v. 145, n. 3, p. 635–640, 1999. DOI: 10.1007/s007050050052.

FULTON, R. W. Practices and Precautions in the Use of Cross Protection for Plant Virus Disease Control. **Annual Review of Phytopathology**, v. 24, n. 1, p. 67–81, 1986. DOI: 10.1146/annurev.py.24.090186.000435.

GAAFAR, Y. Z. A. *et al.* Interlaboratory Comparison Study on Ribodepleted Total RNA High-Throughput Sequencing for Plant Virus Diagnostics and Bioinformatic Competence. **Pathogens**, v. 10, n. 9, p. 1174, 2021. DOI: 10.3390/pathogens10091174.

GABRENAITE-VERKHOVSKAYA, R. *et al.* Cylindrical inclusion protein of Potato virus A is associated with a subpopulation of particles isolated from infected plants. **Journal of General Virology**, v. 89, n. 3, p. 829–838, 2008. DOI: 10.1099/vir.0.83406-0.

GARCÍA, J. A. *et al.* Proteolytic processing of the plum pox potyvirus polyprotein by the Nla protease at a novel cleavage site. **Virology**, v. 188, n. 2, p. 697–703, 1992. DOI: 10.1016/0042-6822(92)90524-S.

GAUT, B. S. Variation in the rates of molecular evolution in plants. **Annual Review of Ecology, Evolution, and Systematics**, v. 42, p. 245–266, 2011. DOI: 10.1146/annurev-ecolsys-102710-145119.

GEMAN, S.; GEMAN, D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. PAMI-6, n. 6, p. 721–741, 1984. DOI: 10.1109/TPAMI.1984.4767596.

GIACHETTI, C.; SEMLER, B. L. Role of a viral membrane polypeptide in strand-specific initiation of poliovirus RNA synthesis. **Journal of Virology**, v. 65, n. 5, p. 2647–2654, 1991. DOI: 10.1128/jvi.65.5.2647-2654.1991.

GIBBS, A. J. *et al.* The prehistory of potyviruses: Their initial radiation was during the dawn of agriculture. **PLoS ONE**, v. 3, n. 6, 2008. DOI: 10.1371/journal.pone.0002523.

GÓMEZ DE CEDRÓN, M. *et al.* Genetic analysis of the function of the plum pox virus CI RNA helicase in virus movement. **Virus Research**, v. 116, n. 1–2, p. 136–145, 2006. DOI: 10.1016/j.virusres.2005.09.009.

GONG, Y. N. *et al.* The Nla-Protease Protein Encoded by the Pepper Mottle Virus Is a Pathogenicity Determinant and Releases DNA Methylation of *Nicotiana benthamiana*. **Frontiers in Microbiology**, v. 11, n. February, p. 1–9, 2020. DOI: 10.3389/fmicb.2020.00102.

GONZÁLEZ-CANDELAS, F.; PATIÑO-GALINDO, J. Á.; VALIENTE-MULLOR, C. Population Genomics of Human Viruses. **Springer Nature**, p. 267–296, 2018. DOI: 10.1007/13836_2018_31.

GONZÁLEZ-JARA, P. *et al.* Host-dependent differences during synergistic infection by Potyviruses with potato virus X. **Molecular Plant Pathology**, v. 5, n. 1, p. 29–35, 2004. DOI: 10.1111/j.1364-3703.2004.00202.x.

GONZÁLEZ RAMÍREZ, J. E. *et al.* Water-Dissolved Ozone Mediates Potyvirus Sanitation during In Vitro Propagation of *Dioscorea Cayenensis* Subsp. Rotundata (Poir.) Miège. **Ozone: Science and Engineering**, v. 42, n. 1, p. 89–94, 2020. DOI: 10.1080/01919512.2019.1598845.

GOTZ, R.; MAISS, E. The complete nucleotide sequence and genome organization of the mite-transmitted brome streak mosaic rymovirus in comparison with those of potyviruses. **Journal of General Virology**, v. 76, n. 8, p. 2035–2042, 1995. DOI: 10.1099/0022-1317-76-8-2035.

GOVIER, D. A.; KASSANIS, B. A virus-induced component of plant sap needed when aphids acquire potato virus Y from purified preparations. **Virology**, v. 61, n. 2, p. 420–426, 1974. DOI: 10.1016/0042-6822(74)90278-5.

HAIKONEN, T. *et al.* Mutation of a short variable region in HCpro protein of Potato virus a affects interactions with a microtubule-associated protein and induces necrotic responses in tobacco. **Molecular Plant-Microbe Interactions**, v. 26, n. 7, p. 721–733, 2013. DOI: 10.1094/MPMI-01-13-0024-R.

HAJIMORAD, M. R.; EGGENBERGER, A. L.; HILL, J. H. Loss and Gain of Elicitor Function of Soybean Mosaic Virus G7 Provoking Rsv1 -Mediated Lethal Systemic Hypersensitive Response Maps to P3. **Journal of Virology**, v. 79, n. 2, p. 1215–1222, 2005. DOI: 10.1128/jvi.79.2.1215-1222.2005.

HALBERT, S. E.; IRWIN, M. E.; GOODMAN, R. M. Alate aphid (Homoptera: Aphididae) species and their relative importance as field vectors of soybean mosaic virus. **Annals of Applied Biology**, v. 97, n. 1, p. 1–9, 1981. DOI: 10.1111/j.1744-7348.1981.tb02988.x.

HASTINGS, W. K. Monte carlo sampling methods using Markov chains and their applications. **Biometrika**, v. 57, n. 1, p. 97–109, 1970. DOI: 10.1093/biomet/57.1.97.

HAYASHI, E. A. I. *et al.* Complete genome sequence of a putative new secovirus infecting yam (*Dioscorea*) plants. **Archives of Virology**, v. 162, n. 1, p. 317–319, 2017. DOI: 10.1007/s00705-016-3104-4.

HELLIOT, B. *et al.* Cryopreservation for the elimination of cucumber mosaic and banana streak viruses from banana (*Musa* spp.). **Plant Cell Reports**, v. 20, n. 12, p. 1117–1122, 2002. DOI: 10.1007/s00299-002-0458-8.

HERNÁNDEZ, J. A. *et al.* Oxidative stress induced by long-term plum pox virus infection in peach (*Prunus persica*). **Physiologia Plantarum**, v. 122, n. 4, p. 486–495, 2004. DOI: 10.1111/j.1399-3054.2004.00431.x.

HERNÁNDEZ, M. *et al.* Bioinformatics of next generation sequencing in clinical microbiology diagnosis. **Revista Argentina de Microbiología**, v. 52, n. 2, p. 150–161, 2020. DOI: 10.1016/j.ram.2019.06.003.

HILY, J. M. *et al.* Plum pox virus coat protein gene intron-hairpin-RNA (ihpRNA) constructs provide resistance to plum pox virus in *Nicotiana benthamiana* and *Prunus domestica*. **Journal of the American Society for Horticultural Science**, v. 132, n. 6, p. 850–858, 2007. DOI: 10.21273/jashs.132.6.850.

HISA, Y. *et al.* P3N-PIPO of Clover yellow vein virus exacerbates symptoms in pea infected with White clover mosaic virus and is implicated in viral synergism. **Virology**, v. 449, p. 200–206, 2014. DOI: 10.1016/j.virol.2013.11.016.

HISLOP, L. *et al.* Sugarcane Mosaic Virus Resistance in the Wisconsin Sweet Corn Diversity Panel. **Journal of the American Society for Horticultural Science**, v. 146, n. 6, p. 435–444, 2021. DOI: 10.21273/jashs05097-21.

HO, S. Y. W. *et al.* The genome as an evolutionary timepiece. **Genome Biology and Evolution**, v. 8, n. 9, p. 3006–3010, 2016. DOI: 10.1093/gbe/evw220.

HO, S. Y. W.; DUCHÊNE, S. Molecular-clock methods for estimating evolutionary rates and timescales. **Molecular Ecology**, v. 23, n. 24, p. 5947–5965, 2014. DOI: 10.1111/mec.12953.

HO, S. Y. W.; JERMIIN, L. S. Tracing the decay of the historical signal in biological sequence data. **Systematic Biology**, v. 53, n. 4, p. 623–637, 2004. DOI: 10.1080/10635150490503035.

HOGENHOUT, S. A. *et al.* Insect vector interactions with persistently transmitted viruses. **Annual Review of Phytopathology**, v. 46, p. 327–359, 2008. DOI: 10.1146/annurev.phyto.022508.092135.

HU, W. *et al.* Complete genomic sequence of a novel macluravirus, alpinia oxyphylla mosaic virus (AloMV), identified in *Alpinia oxyphylla*. **Archives of Virology**, v. 163, n. 9, p. 2579–2582, 2018. DOI: 10.1007/s00705-018-3879-6.

HUELSENBECK, J. P.; RONQUIST, F. MRBAYES: Bayesian inference of phylogenetic trees. **Bioinformatics**, v. 17, n. 8, p. 754–755, 2001. DOI: 10.1093/bioinformatics/17.8.754.

HUET, H. *et al.* Mutations in the helper component protease gene of zucchini yellow mosaic virus affect its ability to mediate aphid transmissibility. **Journal of General Virology**, v. 75, n. 6, p. 1407–1414, 1994. DOI: 10.1099/0022-1317-75-6-1407.

HUTTLEY, G. A.; WAKEFIELD, M. J.; EASTEAL, S. Rates of genome evolution and branching order from whole genome analysis. **Molecular Biology and Evolution**, v. 24, n. 8, p. 1722–1730, 2007. DOI: 10.1093/molbev/msm094.

IDRIS, A. *et al.* Viral metagenomics: Analysis of begomoviruses by illumina high-throughput sequencing. **Viruses**, v. 6, n. 3, p. 1219–1236, 2014. DOI: 10.3390/v6031219.

INOUE-NAGATA, A. K. *et al.* A simple method for cloning the complete begomovirus genome using the bacteriophage ϕ 29 DNA polymerase. **Journal of Virological Methods**, v. 116, n. 2, p. 209–211, 2004. DOI: 10.1016/j.jviromet.2003.11.015.

ITA, E. E. *et al.* Efficient elimination of Yam mosaic virus (YMV) from white yam (*Dioscorea rotundata* Poir.) by cryotherapy of axillary buds. **South African Journal of Botany**, v. 130, p. 123–129, 2020. DOI: 10.1016/j.sajb.2019.12.022.

JANSSEN, D. *et al.* Absence of a coding region for the helper component-proteinase in the genome of cucumber vein yellowing virus, a whitefly-transmitted member of the Potyviridae. **Archives of Virology**, v. 150, n. 7, p. 1439–1447, 2005. DOI: 10.1007/s00705-005-0515-z.

JARUGULA, S. *et al.* First Report of grapevine red globe virus in grapevines in Washington State. **Plant Disease**, p. 717, 2021. DOI: 10.1094/PDIS-07-20-1609-PDN.

JENNER, C. E. *et al.* dual role of the potyvirus P3 protein of Turnip mosaic virus as a symptom and avirulence determinant in brassicas. **Molecular Plant-Microbe Interactions**, v. 16, n. 9, p. 777–784, 2003. DOI: 10.1094/MPMI.2003.16.9.777.

JIANG, C. *et al.* Bymovirus-induced yellow mosaic diseases in barley and wheat: viruses, genetic resistances and functional aspects. **Theoretical and Applied Genetics**, v. 133, n. 5, p. 1623–1640, 2020. DOI: 10.1007/s00122-020-03555-7.

JO, Y.; CHOI, H.; CHO, W. K. De novo genome assembly of ryegrass mosaic virus from a ryegrass transcriptome. **Genome Announcements**, v. 3, n. 3, p. 6–7, 2015. DOI: 10.1128/genomeA.00497-15.

JOHANSEN, I. E. *et al.* Recessive Resistance in *Pisum sativum* and Potyvirus Pathotype Resolved in a Gene-for-Cistron Correspondence between Host and Virus. **Journal of Virology**, v. 75, n. 14, p. 6609–6614, 2001. DOI: 10.1128/jvi.75.14.6609-6614.2001.

JOHNSON, K. L.; SARNOV, P. Three poliovirus 2B mutants exhibit noncomplementable defects in viral RNA amplification and display dosage-dependent dominance over wild-type poliovirus. **Journal of Virology**, v. 65, n. 8, p. 4341–4349, 1991. DOI: 10.1128/jvi.65.8.4341-4349.1991.

JONES, R. A. C. Strain group specific and virus specific hypersensitive reactions to infection with potyviruses in potato cultivars. **Annals of Applied Biology**, v. 117, n. 1, p. 93–105, 1990. DOI: 10.1111/j.1744-7348.1990.tb04198.x.

JONES, R. A. C. Strain-Specific Hypersensitive and Extreme Resistance Phenotypes Elicited by Potato virus Y Among 39 Potato Cultivars Released in Three World Regions Over a 117-Year Period. **Plant Disease**, p. 185–196, 2018. DOI: 10.1094/PDIS-06-17-0901-RE.

JOSSEY, S.; HOBBS, H. A.; DOMIER, L. L. Role of soybean mosaic virus-encoded proteins in seed and aphid transmission in soybean. **Phytopathology**, v. 103, n. 9, p. 941–948, 2013. DOI: 10.1094/PHYTO-09-12-0248-R.

JOVER-GIL, S. *et al.* Complete genome sequence of a novel virus, classifiable within the Potyviridae family, which infects passion fruit (*Passiflora edulis*). **Archives of Virology**, v. 163, n. 11, p. 3191–3194, 2018. DOI: 10.1007/s00705-018-3983-7.

KALLESHWARASWAMY, C. M.; KRISHNA KUMAR, N. K. Transmission efficiency of Papaya ringspot virus by three aphid species. **Phytopathology**, v. 98, n. 5, p. 541–546, 2008. DOI: 10.1094/PHYTO-98-5-0541.

KANNAN, M.; ISMAIL, I.; BUNAWAN, H. Maize dwarf mosaic virus: From genome to disease management. **Viruses**, v. 10, n. 9, p. 1–23, 2018. DOI: 10.3390/v10090492.

KASSCHAU, K. D.; CARRINGTON, J. C. A counterdefensive strategy of plant viruses: Suppression of posttranscriptional gene silencing. **Cell**, v. 95, n. 4, p. 461–470, 1998. DOI: 10.1016/S0092-8674(00)81614-1.

KASSCHAU, K. D. *et al.* P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA function. **Developmental Cell**, v. 4, n. 2, p. 205–217, 2003. DOI: 10.1016/S1534-5807(03)00025-X.

KEKARAINEN, T. *et al.* Comparison of the complete sequences of five different isolates of Potato virus A (PVA), genus Potyvirus. **Archives of Virology**, v. 144, n. 12, p. 2355–2366, 1999. DOI: 10.1007/s007050050649.

KEKARAINEN, T.; SAVILAHTI, H.; VALKONEN, J. Functional genomics on Potato virus A: Virus genome-wide map of sites essential for virus propagation. **Genome Research**, v. 12, n. 4, p. 584–594, 2002. DOI: 10.1101/gr.220702.

KELLY, D. L.; SMITH, C. L. Bayesian inference in probabilistic risk assessment-The current state of the art. **Reliability Engineering and System Safety**, v. 94, n. 2, p. 628–643, 2009. DOI: 10.1016/j.ress.2008.07.002.

KELMANSON, J. E.; JA, A. K.; STADEN, J. V. Zulu medicinal plants with antibacterial activity. **Journal of Ethnopharmacology**, v. 69, p. 241–246, 2000.

KENYON, L. *et al.* An overview of viruses infecting Dioscorea yams in sub-Saharan Africa. **Plant Virology in Sub-Saharan Africa**, n. May 2014, p. 432–439, 2001.

KERSCH-BECKER, M. F.; THALER, J. S. Virus strains differentially induce plant susceptibility to aphid vectors and chewing herbivores. **Oecologia**, v. 174, n. 3, p. 883–892, 2014. DOI: 10.1007/s00442-013-2812-7.

KIM, B. M. *et al.* The P3 protein of turnip mosaic virus can alone induce hypersensitive response-like cell death in Arabidopsis thaliana carrying TuNI. **Molecular Plant-Microbe Interactions**, v. 23, n. 2, p. 144–152, 2010. DOI: 10.1094/MPMI-23-2-0144.

KIM, B. *et al.* Veinal necrosis induced by Turnip mosaic virus infection in arabidopsis is a form of defense response accompanying HR-like cell death. **Molecular Plant-Microbe Interactions**, v. 21, n. 2, p. 260–268, 2008. DOI: 10.1094/MPMI-21-2-0260.

KIM, S. B. *et al.* RNA-dependent RNA polymerase (NIb) of the potyviruses is an avirulence factor for the broad-spectrum resistance gene Pvr4 in Capsicum annuum cv. CM334. **PLoS ONE**, v. 10, n. 3, p. 1–11, 2015. DOI: 10.1371/journal.pone.0119639.

KIMURA, M. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. **Genetics**, v. 61, n. 4, p. 893–903, 1969. DOI: 10.1093/genetics/61.4.893.

KIMURA, M. Rare variant alleles in the light of the neutral theory. **Molecular Biology and Evolution**, v. 1, n. 1, p. 84–93, 1983. DOI: 10.1093/oxfordjournals.molbev.a040305.

KIMURA, M.; OTA, T. On some principles governing molecular evolution. **Proceedings of the National Academy of Sciences of the United States of America**, v. 71, n. 7, p. 2848–2852, 1974. DOI: 10.1073/pnas.71.7.2848.

KIMURA, M. Evolutionary Rate at the Molecular Level. **Nature**, n. 217, p. 624–626, 1968. DOI: <https://doi.org/10.1038/217624a0>.

KIMURA, M. Molecular Evolutionary Clock and the Neutral Theory. **Journal of Molecular Evolution**, v. 26, p. 24–33, 1987. DOI: 10.1007/BF02111279.

KIRAN R., G. *et al.* Aphid Transmission of Potyvirus: The Largest Plant-Infecting RNA Virus Genus. **Viruses**, v. 12, n. 7, p. 773, 2020. DOI: 10.3390/v12070773.

KITAJIMA, E. W. An annotated list of plant viruses and viroids described in Brazil (1926–2018). **Biota Neotropica**, v. 20, n. 2, 2020. DOI: 10.1590/1676-0611-bn-2019-0932.

KONDO, T.; FUJITA, T. Complete nucleotide sequence and construction of an infectious clone of Chinese yam necrotic mosaic virus suggest that macluraviruses have the smallest genome among members of the family Potyviridae. **Archives of Virology**, v. 157, n. 12, p. 2299–2307, 2012. DOI: 10.1007/s00705-012-1429-1.

KOSAKA, Y.; FUKUNISHI, T. Multiple inoculation with three attenuated viruses for the control of cucumber virus disease. **Plant Disease**, v. 81, n. 7, p. 733–738, 1997. DOI: 10.1094/PDIS.1997.81.7.733.

KOSAKA Y, F. T. Attenuated Isolates of Soybean Mosaic Virus Derived at a Low Temperature. **Plant Disease**, v. 77, p. 882–886, 1993. DOI: 10.1094/PD-77-0882.

KOSUGI, S.; HIRAKAWA, H.; TABATA, S. GMcloser: Closing gaps in assemblies accurately with a likelihood-based selection of contig or long-read alignments. **Bioinformatics**, v. 31, n. 23, p. 3733–3741, 2015. DOI: 10.1093/bioinformatics/btv465.

KREUZE, J. F. *et al.* Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. **Virology**, v. 388, n. 1, p. 1–7, 2009. DOI: 10.1016/j.virol.2009.03.024.

KUMAR, S. Molecular clocks: Four decades of evolution. **Nature Reviews Genetics**, v. 6, n. 8, p. 654–662, 2005. DOI: 10.1038/nrg1659.

KUMAR, S; BLAIR HEDGES, S. Advances in time estimation methods for molecular data. **Molecular Biology and Evolution**, v. 33, n. 4, p. 863–869, 2016. DOI: 10.1093/molbev/msw026.

KURIHARA, Y.; WATANABE, Y. Cross-protection in *Arabidopsis* against crucifer tobamovirus Cg by an attenuated strain of the virus. **Molecular Plant Pathology**, v. 4, n. 4, p. 259–269, 2003. DOI: 10.1046/j.1364-3703.2003.00174.x.

LAI, B. *et al.* InteMAP: Integrated metagenomic assembly pipeline for NGS short reads. **BMC Bioinformatics**, v. 16, n. 1, p. 1–14, 2015. DOI: 10.1186/s12859-015-0686-x.

LAN, P. *et al.* Complete genome sequence of yam chlorotic necrosis virus, a novel macluravirus infecting yam. **Archives of Virology**, v. 163, n. 8, p. 2275–2278, 2018. DOI: 10.1007/s00705-018-3851-5.

LANDER, E. S. *et al.* Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860–921, 2001. DOI: 10.1038/35057062.

LARGET, B.; SIMON, D. L. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. **Molecular Biology and Evolution**, v. 16, n. 6, p. 750–759, 1999. DOI: 10.1093/oxfordjournals.molbev.a026160.

LARTILLOT, N.; PHILIPPE, H. Computing Bayes factors using thermodynamic integration. **Systematic Biology**, v. 55, n. 2, p. 195–207, 2006. DOI: 10.1080/10635150500433722.

LEFEUVRE, P. *et al.* Evolution and ecology of plant viruses. **Nature Reviews Microbiology**, v. 17, n. 10, p. 632–644, 2019. DOI: 10.1038/s41579-019-0232-3.

LEPAGE, T. *et al.* A general comparison of relaxed molecular clock models. **Molecular Biology and Evolution**, v. 24, n. 12, p. 2669–2680, 2007. DOI: 10.1093/molbev/msm193.

LEVY, S. E.; BOONE, B. E. Next-generation sequencing strategies. **Cold Spring Harbor Perspectives in Medicine**, v. 9, n. 7, p. 1–11, 2019. DOI: 10.1101/cshperspect.a025791.

LEVY, S. E.; MYERS, R. M. Advancements in Next-Generation Sequencing. **Annual Review of Genomics and Human Genetics**, v. 17, n. May, p. 95–115, 2016. DOI: 10.1146/annurev-genom-083115-022413.

LI, D. *et al.* MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. **Bioinformatics**, v. 31, n. 10, p. 1674–1676, 2015. DOI: 10.1093/bioinformatics/btv033.

LI, F. *et al.* Phylogenetic relationships of closely related potyviruses infecting sweet potato determined by genomic characterization of Sweet potato virus G and Sweet potato virus 2. **Virus Genes**, v. 45, n. 1, p. 118–125, 2012. DOI: 10.1007/s11262-012-0749-2.

LI, W. *et al.* Presence of P1b and absence of HC-Pro in Squash vein yellowing virus suggests a general feature of the genus *Ipomovirus* in the family *Potyviridae*. **Virus Research**, v. 135, n. 2, p. 213–219, 2008. DOI: 10.1016/j.virusres.2008.03.015.

LI, W. H.; WU, CI; LUO, C. C. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. **Mol Biol Evol.**, v. 2, n. 2, p. 150–174, 1985. DOI: 10.1093/oxfordjournals.molbev.a040343.

LIN, J. An integrated procedure for bayesian reliability inference using MCMC. **Journal of Quality and Reliability Engineering**, v. 2014, 2014. DOI: 10.1155/2014/264920.

LIN, J.; ASPLUND, M. Bayesian semi-parametric analysis for locomotive wheel degradation using gamma frailties. **Rail and Rapid Transit**, v. 229, n. 3, p. 237–247, 2015. DOI: 10.1177/0954409713508759.

LIN, J.; ASPLUNDA, M.; PARIDAA, A. Reliability analysis for degradation of locomotive wheels using parametric bayesian approach. **Quality and Reliability Engineering International**, v. 30, n. 5, p. 657–667, 2014. DOI: 10.1002/qre.1518.

LIU, X. *et al.* Cloning and transformation of SCMV CP gene and regeneration of transgenic maize plants showing resistance to SCMV strain MDB. **African Journal of Biotechnology**, v. 8, n. 16, p. 3747–3753, 2009. DOI: 10.4314/ajb.v8i16.61926.

LÓPEZ-MOYA, J. J.; WANG, R. Y.; PIRONE, T. P. Context of the coat protein DAG motif affects potyvirus transmissibility by aphids. **Journal of General Virology**, v. 80, n. 12, p. 3281–3288, 1999. DOI: 10.1099/0022-1317-80-12-3281.

LU, H.; GIORDANO, F.; NING, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. **Genomics, Proteomics and Bioinformatics**, v. 14, n. 5, p. 265–279, 2016. DOI: 10.1016/j.gpb.2016.05.004.

LUO, A.; HO, S. Y. W. The molecular clock and evolutionary timescales. **Biochemical Society Transactions**, v. 46, n. 5, p. 1183–1190, 2018. DOI: 10.1042/BST20180186.

LUO, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. **GigaScience**, v. 4, n. 1, p. 1–18, 2012. DOI: 10.1186/2047-217X-1-18.

MACMANES, M. D. On the optimal trimming of high-throughput mRNA sequence data. **Frontiers in Genetics**, v. 5, n. JAN, p. 1–7, 2014. DOI: 10.3389/fgene.2014.00013.

MALAPA, R. *et al.* Genetic diversity of the greater yam (*Dioscorea alata* L.) and relatedness to *D. nummularia* Lam. and *D. transversa* Br. as revealed with AFLP markers. **Genetic Resources and Crop Evolution**, v. 52, n. 7, p. 919–929, 2005. DOI: 10.1007/s10722-003-6122-5.

MALIOGKA, V. I. *et al.* Virus variants with differences in the p1 protein coexist in a plum pox virus population and display particular host-dependent pathogenicity features. **Molecular Plant Pathology**, v. 13, n. 8, p. 877–886, 2012. DOI: 10.1111/j.1364-3703.2012.00796.x.

MALIOGKA, V. I. *et al.* Recent advances on detection and characterization of fruit tree viruses using high-throughput sequencing technologies. **Viruses**, v. 10, n. 8, p. 1–23, 2018. DOI: 10.3390/v10080436.

MALLORY, A. C. *et al.* A viral suppressor of RNA silencing differentially regulates the accumulation of short interfering RNAs and micro-RNAs in tobacco. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 23, p. 15228–15233, 2002. DOI: 10.1073/pnas.232434999.

MAMBOLE, I. A. *et al.* Molecular characterization of yam virus X, a new potexvirus infecting yams (*Dioscorea* spp) and evidence for the existence of at least three distinct potexviruses infecting yams. **Archives of Virology**, v. 159, n. 12, p. 3421–3426, 2014. DOI: 10.1007/s00705-014-2211-3.

MARAIS, A. *et al.* Yam asymptomatic virus 1, a novel virus infecting yams (*Dioscorea* spp.) with significant prevalence in a germplasm collection. **Archives of Virology**, v. 165, n. 11, p. 2653–2657, 2020. DOI: 10.1007/s00705-020-04787-0.

MARTIN, B. *et al.* Intracellular ingestion and salivation by aphids may cause the acquisition and inoculation of non-persistently transmitted plant viruses. **Journal of General Virology**, v. 78, n. 10, p. 2701–2705, 1997. DOI: 10.1099/0022-1317-78-10-2701.

MARTINEZ, F.; DAROS, J. A. Tobacco etch virus Protein P1 Traffics to the Nucleolus and Associates with the Host 60S Ribosomal Subunits during Infection. **Journal of Virology**, v. 88, n. 18, p. 10725–10737, 2014. DOI: 10.1128/jvi.00928-14.

MARTÍNEZ, F. *et al.* Interaction network of tobacco etch potyvirus NIa protein with the host proteome during infection. **BMC Genomics**, v. 17, n. 1, p. 1–13, 2016. DOI: 10.1186/s12864-016-2394-y.

MAULE, A. J. Virus and Host Plant Interactions. **eLS**, n. D, 2001. DOI: 10.1038/npg.els.0000758.

MBANZIBWA, D. R. *et al.* Simultaneous virus-specific detection of the two cassava brown streak-associated viruses by RT-PCR reveals wide distribution in East Africa, mixed infections, and infections in *Manihot glaziovii*. **Journal of Virological Methods**, v. 171, n. 2, p. 394–400, 2011. DOI: 10.1016/j.jviromet.2010.09.024.

MCMULLEN, M. D. Identification of a Gene for Resistance to Wheat Streak Mosaic Virus in Maize. **Phytopathology**, 1991. DOI: 10.1094/phyto-81-624.

MEDVEDEV, P. *et al.* Paired de Bruijn graphs: A novel approach for incorporating mate pair information into genome assemblers. **Journal of Computational Biology**, v. 18, n. 11, p. 1625–1634, 2011. DOI: 10.1089/cmb.2011.0151.

MELCHER, U. *et al.* Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from *Ambrosia psilostachya*. **Journal of Virological Methods**, v. 152, n. 1–2, p. 49–55, 2008. DOI: 10.1016/j.jviromet.2008.05.030.

MENZEL, W.; THOTTAPPILLY, G.; WINTER, S. Characterization of an isometric virus isolated from yam (*Dioscorea rotundata*) in Nigeria suggests that it belongs to a new species in the genus *Aureusvirus*. **Archives of Virology**, v. 159, n. 3, p. 603–606, 2014. DOI: 10.1007/s00705-013-1845-x.

METROPOLIS, N. *et al.* Equation of state calculations by fast computing machines. **The Journal of Chemical Physics**, v. 21, n. 6, p. 1087–1092, 1953. DOI: 10.1063/1.1699114.

MILLER, J. R. *et al.* Aggressive assembly of pyrosequencing reads with mates. **Bioinformatics**, v. 24, n. 24, p. 2818–2824, 2008. DOI: 10.1093/bioinformatics/btn548.

MILLER, J. R.; KOREN, S.; SUTTON, G. Assembly Algorithms for Next-Generation Sequencing Data. **Genomics**, v. 95, n. 6, p. 315–327, 2010. DOI: 10.1016/j.ygeno.2010.03.001.

MINEI, R.; HOSHINA, R.; OGURA, A. De novo assembly of middle-sized genome using MinION and Illumina sequencers 06 Biological Sciences 0604 Genetics. **BMC Genomics**, v. 19, n. 1, p. 1–12, 2018. DOI: 10.1186/s12864-018-5067-1.

MINGOT, A. *et al.* The P1N-PISPO trans -Frame Gene of Sweet Potato Feathery Mottle Potyvirus Is Produced during Virus Infection and Functions as an RNA Silencing Suppressor. **Journal of Virology**, v. 90, n. 7, p. 3543–3557, 2016. DOI: 10.1128/jvi.02360-15.

MINUTILLO, S. A. *et al.* Complete nucleotide sequence of Artichoke latent virus shows it to be a member of the genus *Macluravirus* in the family *Potyviridae*. **Phytopathology**, v. 105, n. 8, p. 1155–1160, 2015. DOI: 10.1094/PHYTO-01-15-0010-R.

MODI, A. *et al.* The Illumina Sequencing Protocol and the NovaSeq 6000 System. **Methods in Molecular Biology**, v. 2242, p. 15–42, 2021. DOI: 10.1007/978-1-0716-1099-2_2.

MOLLOV, D.; LOCKHART, B.; ZLESK, D. Complete nucleotide sequence of rose yellow mosaic virus, a novel member of the family *Potyviridae*. **Archives of Virology**, v. 158, n. 9, p. 1917–1923, 2013. DOI: 10.1007/s00705-013-1686-7.

MOORJANI, P. *et al.* Variation in the molecular clock of primates. **Proceedings of the National Academy of Sciences of the United States of America**, v. 113, n. 38, p. 10607–10612, 2016. DOI: 10.1073/pnas.1600374113.

MORENO, A. *et al.* Differences in the mechanism of inoculation between a semi-persistent and a non-persistent aphid-transmitted plant virus. **Journal of General Virology**, v. 93, n. 3, p. 662–667, 2012. DOI: 10.1099/vir.0.037887-0.

MORGAN, G. J. Emile Zuckerkandl, Linus Pauling, and the Molecular Evolutionary Clock, 1959-1965. **Journal of the History of Biology**, v. 31, p. 155–178, 1998. Disponível em: <https://www.jstor.org/stable/4331476>.

MURPHY, J. F. *et al.* Replacement of the tyrosine residue that links a potyviral VPg to the viral RNA is lethal. **Virology**, v. 220, n. 2, p. 535–538, 1996. DOI: 10.1006/viro.1996.0344.

MURRY, L. E. *et al.* Transgenic Corn Plants Expressing MDMV Strain B Coat Protein are Resistant to Mixed Infections of Maize Dwarf Mosaic Virus and Maize Chlorotic Mottle Virus. **Nat Biotechnol**, v. 11, p. 1559–1564, 1993. DOI: <https://doi.org/10.1038/nbt1293-1559>.

MUTHUKUMAR, V. *et al.* Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. **Virus Research**, v. 141, n. 2, p. 169–173, 2009. DOI: 10.1016/j.virusres.2008.06.016.

NAGARAJAN, N.; POP, M. Sequence assembly demystified. **Nature Reviews Genetics**, v. 14, n. 3, p. 157–167, 2013. DOI: 10.1038/nrg3367.

NAKAZONO-NAGAOKA, E. *et al.* Cross-protection against bean yellow mosaic virus (BYMV) and clover yellow vein virus by attenuated BYMV Isolate M11. **Phytopathology**, v. 99, n. 3, p. 251–257, 2009. DOI: 10.1094/PHYTO-99-3-0251.

NAMIKI, T. *et al.* MetaVelvet: An extension of Velvet assembler to de novo metagenome assembly from short sequence reads. **Nucleic Acids Research**, v. 40, n. 20, 2012. DOI: 10.1093/nar/gks678.

NAVARRO, B. *et al.* The first phlebo-like virus infecting plants: a case study on the adaptation of negative-stranded RNA viruses to new hosts. **Molecular Plant Pathology**, v. 19, n. 5, p. 1075–1089, 2018. DOI: 10.1111/mpp.12587.

NG, J. C. K.; PERRY, K. L. Transmission of plant viruses by aphid vectors. **Molecular Plant Pathology**, v. 5, n. 5, p. 505–511, 2004. DOI: 10.1111/J.1364-3703.2004.00240.X.

NGWE, M. F. S. N.; OMOKOLO, D. N.; JOLY, S. Evolution and Phylogenetic Diversity of Yam Species (*Dioscorea* spp.): Implication for Conservation and Agricultural Practices. **PLoS ONE**, v. 10, n. 12, 2015. DOI: 10.1371/journal.pone.0145364.

NIEPEL, M.; GALLIE, D. R. Identification and Characterization of the Functional Elements within the Tobacco Etch Virus 5' Leader Required for Cap-Independent Translation. **Journal of Virology**, v. 73, n. 11, p. 9080–9088, 1999. DOI: 10.1128/jvi.73.11.9080-9088.1999.

NIGAM, D. *et al.* Genome-Wide Variation in Potyviruses. **Frontiers in Plant Science**, v. 10, n. November, p. 1–28, 2019. DOI: 10.3389/fpls.2019.01439.

OLMOS, A. *et al.* High-throughput sequencing technologies for plant pest diagnosis: challenges and opportunities. **EPPO Bulletin**, v. 48, n. 2, p. 219–224, 2018. DOI: 10.1111/epp.12472.

- OLSPERT, A. *et al.* Transcriptional slippage in the positive-sense RNA virus family Potyviridae. **EMBO reports**, v. 16, n. 8, p. 995–1004, 2015. DOI: 10.15252/embr.201540509.
- PASIN, F.; SIMÓN-MATEO, C.; GARCÍA, J. A. The Hypervariable Amino-Terminus of P1 Protease Modulates Potyviral Replication and Host Defense Responses. **PLoS Pathogens**, v. 10, n. 3, 2014. DOI: 10.1371/journal.ppat.1003985.
- PASZKIEWICZ, K.; STUDHOLME, D. J. De novo assembly of short sequence reads. **Briefings in Bioinformatics**, v. 11, n. 5, p. 457–472, 2010. DOI: 10.1093/bib/bbq020.
- PAULINO, D. *et al.* Sealer: A scalable gap-closing application for finishing draft genomes. **BMC Bioinformatics**, v. 16, n. 1, p. 1–8, 2015. DOI: 10.1186/s12859-015-0663-4.
- PEERENBOOM, E. *et al.* Complete RNA1 sequences of two UK isolates of barley mild mosaic virus: A wild-type fungus-transmissible isolate and a non-fungus-transmissible derivative. **Virus Research**, v. 50, n. 2, p. 175–183, 1997. DOI: 10.1016/S0168-1702(97)00069-5.
- PENG, Y. H. *et al.* Mutations in the HC-Pro gene of zucchini yellow mosaic potyvirus: Effects on aphid transmission and binding to purified virions. **Journal of General Virology**, v. 79, n. 4, p. 897–904, 1998. DOI: 10.1099/0022-1317-79-4-897.
- PENG, Y. *et al.* IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. **Bioinformatics**, v. 28, n. 11, p. 1420–1428, 2012. DOI: 10.1093/bioinformatics/bts174.
- PEREZ, P. *et al.* Estimation of vector propensity of potato virus Y in open-field pepper crops of central Spain. **Journal of Economic Entomology**, v. 88, n. 4, p. 986–991, 1995. DOI: 10.1093/jee/88.4.986.
- PIO-RIBEIRO, G. *et al.* Plantas de inhame em Pernambuco e Paraíba apresentam infecções simples e mista por potyvírus e badnavírus. **Fitopatologia Brasileira**, v. v. 31, p. 309, 2006.
- PIRONE, T.; PERRY, K. L. Aphids: Non-persistent Transmission. **advances in botanical research**, v. 36, p. 1–19, 2002. DOI: doi:10.1016/s0065-2296(02)36056-7.
- POMPE-NOVAK, M.; WRISCHER, M.; RAVNIKAR, M. Ultrastructure of chloroplasts in leaves of potato plants infected by potato virus YNTN. **Phyton - Annales Rei Botanicae**, v. 41, n. 2, p. 215–226, 2001.
- POOGGIN, M. M. Small RNA-omics for plant virus identification, virome reconstruction, and antiviral defense characterization. **Frontiers in Microbiology**, v. 9, n. NOV, p. 1–20, 2018. DOI: 10.3389/fmicb.2018.02779.
- POWELL, G. Intracellular salivation is the aphid activity associated with inoculation or non-persistently transmitted viruses. **Journal of General Virology**, v. 86, n. 2, p. 469–472, 2005. DOI: 10.1099/vir.0.80632-0.

POWELL, G.; PIRONE, T.; HARDIE, J. Aphid stylet activities during potyvirus acquisition from plants and an in vitro system that correlate with subsequent transmission. **European Journal of Plant Pathology**, v. 101, n. 4, p. 411–420, 1995. DOI: 10.1007/BF01874855.

POWELL, G.; TOSH, C. R.; HARDIE, J. Host plant selection by aphids: Behavioral, evolutionary, and applied perspectives. **Annual Review of Entomology**, v. 51, n. 22, p. 309–330, 2006. DOI: 10.1146/annurev.ento.51.110104.151107.

PRAVEEN, S. *et al.* Silencing potential of viral derived RNAi constructs in Tomato leaf curl virus-AC4 gene suppression in tomato. **Transgenic Research**, v. 19, n. 1, p. 45–55, 2010. DOI: 10.1007/s11248-009-9291-y.

PRINS, M. *et al.* Strategies for antiviral resistance in transgenic plants. **Molecular Plant Pathology**, v. 9, n. 1, p. 73–83, 2008. DOI: 10.1111/j.1364-3703.2007.00447.x.

PRJIBELSKI, A. *et al.* Using SPAdes De Novo Assembler. **Current Protocols in Bioinformatics**, v. 70, n. 1, p. 1–29, 2020. DOI: 10.1002/cpbi.102.

PRUSS, G. *et al.* Plant viral synergism: The potyviral genome encodes a broad-range pathogenicity enhancer that transactivates replication of heterologous viruses. **Plant Cell**, v. 9, n. 6, p. 859–868, 1997. DOI: 10.1105/tpc.9.6.859.

PRUSS, G. J. *et al.* The potyviral suppressor of RNA silencing confers enhanced resistance to multiple pathogens. **Virology**, v. 320, n. 1, p. 107–120, 2004. DOI: 10.1016/j.virol.2003.11.027.

RABELO-FILHO, F. A. C. *et al.* The complete genome sequence of a Brazilian isolate of yam mild mosaic virus. **Archives of Virology**, v. 158, n. 2, p. 515–518, 2013. DOI: 10.1007/s00705-012-1509-2.

RAJAMÄKI, M. L. *et al.* A novel insertion site inside the potyvirus P1 cistron allows expression of heterologous proteins and suggests some P1 functions. **Virology**, v. 342, n. 1, p. 88–101, 2005. DOI: 10.1016/j.virol.2005.07.019.

RANNALA, B.; YANG, Z. Inferring speciation times under an episodic molecular clock. **Systematic Biology**, v. 56, n. 3, p. 453–466, 2007. DOI: 10.1080/10635150701420643.

REDINBAUGH, M. G.; ZAMBRANO, J. L. Control of Virus Diseases in Maize. **Advances in Virus Research**, v. 90, p. 391–429, 2014. DOI: 10.1016/B978-0-12-801246-8.00008-1.

RESTREPO, M. A.; FREED, D. D.; CARRINGTON, J. C. Nuclear Transport of Plant Potyviral Proteins. **The Plant Cell**, v. 2, n. 10, p. 987–998, 1990. DOI: 10.1105/tpc.2.10.987.

REUTER, J. A.; SPACEK, D. V.; SNYDER, M. P. High-Throughput Sequencing Technologies. **Molecular Cell**, v. 58, n. 4, p. 586–597, 2015. DOI: 10.1016/j.molcel.2015.05.004.

REVERS, F.; GARCÍA, J. A. Molecular biology of potyviruses. **Advances in Virus Research**, v. 92, p. 101–199, 2015. DOI: 10.1016/bs.aivir.2014.11.006.

RHOADS, A.; AU, K. F. PacBio Sequencing and Its Applications. **Genomics, Proteomics and Bioinformatics**, v. 13, n. 5, p. 278–289, 2015. DOI: 10.1016/j.gpb.2015.08.002.

RIECHMANN, J. L.; CERVERA, M. T.; GARCIA, J. A. Processing of the plum pox virus polyprotein at the P3-6K1 junction is not required for virus viability. **Journal of General Virology**, v. 76, n. 4, p. 951–956, 1995. DOI: 10.1099/0022-1317-76-4-951.

RODAMILANS, B.; CASILLAS, A.; GARCÍA, J. A. P1 of Sweet Potato Feathery Mottle Virus Shows Strong Adaptation Capacity, Replacing P1-HCPro in a Chimeric Plum Pox Virus. **Journal of Virology**, v. 95, n. 14, 2021. DOI: 10.1128/jvi.00150-21.

RODAMILANS, B. *et al.* RNA Polymerase Slippage as a Mechanism for the Production of Frameshift Gene Products in Plant Viruses of the Potyviridae Family. **Journal of Virology**, v. 89, n. 13, p. 6965–6967, 2015. DOI: 10.1128/jvi.00337-15.

RODAMILANS, B. *et al.* An atypical RNA silencing suppression strategy provides a snapshot of the evolution of sweet potato-infecting potyviruses. **Scientific Reports**, v. 8, n. 1, p. 1–10, 2018. DOI: 10.1038/s41598-018-34358-y.

RODRÍGUEZ-CEREZO, E.; SHAW, J. G. Two newly detected nonstructural viral proteins in potyvirus-infected cells. **Virology**, v. 185, n. 2, p. 572–579, 1991. DOI: 10.1016/0042-6822(91)90527-I.

RODRÍGUEZ, M. *et al.* The chlorotic symptom induced by Sunflower chlorotic mottle virus is associated with changes in redox-related gene expression and metabolites. **Plant Science**, v. 196, p. 107–116, 2012. DOI: 10.1016/j.plantsci.2012.08.008.

RODRÍGUEZ, M. *et al.* Are Sunflower chlorotic mottle virus infection symptoms modulated by early increases in leaf sugar concentration? **Journal of Plant Physiology**, v. 167, n. 14, p. 1137–1144, 2010. DOI: 10.1016/j.jplph.2010.03.004.

ROOSSINCK, M. J.; MARTIN, D. P.; ROUMAGNAC, P. Plant virus metagenomics: Advances in virus discovery. **Phytopathology**, v. 105, n. 6, p. 716–727, 2015. DOI: 10.1094/PHYTO-12-14-0356-RVW.

ROSARIO, K. *et al.* Discovery of a novel mastrevirus and alphasatellite-like circular DNA in dragonflies (Ephemeroptera) from Puerto Rico. **Virus Research**, v. 171, n. 1, p. 231–237, 2013. DOI: 10.1016/j.virusres.2012.10.017.

ROSE, H. *et al.* Complete genome sequence of a German isolate of spartina mottle virus supports its classification as a member of the proposed genus “Sparmovirus” within the family Potyviridae. **Archives of Virology**, v. 165, n. 10, p. 2385–2388, 2020. DOI: 10.1007/s00705-020-04726-z.

ROSE, H. *et al.* Complete genome sequence and construction of an infectious full-length cDNA clone of celery latent virus – An unusual member of a putative new genus within the potyviridae. **Journal of General Virology**, v. 100, n. 2, p. 308–320, 2019. DOI: 10.1099/jgv.0.001207.

ROTHBERG, J. M. *et al.* An integrated semiconductor device enabling non-optical genome sequencing. **Nature**, v. 475, n. 7356, p. 348–352, 2011. DOI: 10.1038/nature10242.

ROTHBERG, J. M.; LEAMON, J. H. The development and impact of 454 sequencing. **Nature Biotechnology**, v. 26, n. 10, p. 1117–1124, 2008. DOI: 10.1038/nbt1485.

SÁENZ, P. *et al.* Identification of a pathogenicity determinant of Plum pox virus in the sequence encoding the C-terminal region of protein P3+6K1. **Journal of General Virology**, v. 81, n. 3, p. 557–566, 2000. DOI: 10.1099/0022-1317-81-3-557.

SALEIL, V. *et al.* Obtention de plantes indemnes du virus de la mosaïque de l'igname (YMV) par culture in vitro des apex chez l'igname américaine *Dioscorea trifida* L. **Agronomie**, v. 10, n. 8, p. 605–615, 1990. DOI: 10.1051/agro:19900801.

SANDERSON, M. J. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. **Molecular Biology and Evolution**, v. 19, n. 1, p. 101–109, 2002. DOI: 10.1093/oxfordjournals.molbev.a003974.

SANGER, F.; COULSON, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. **Journal of Molecular Biology**, v. 94, n. 3, p. 441–448, 1975. DOI: 10.1016/0022-2836(75)90213-2.

SANJUÁN, R. *et al.* Viral Mutation Rates. **Journal of Virology**, v. 84, n. 19, p. 9733–9748, 2010. DOI: 10.1128/jvi.00694-10.

SCHAAD, M. C.; JENSEN, P. E.; CARRINGTON, J. C. Formation of plant RNA virus replication complexes on membranes: role of an endoplasmic reticulum-targeted viral protein. **The EMBO journal**, v. 16, n. 13, p. 4049–4059, 1997. DOI: 10.1093/emboj/16.13.4049.

SCOTT, G. E. Linkage between maize dwarf mosaic virus resistance and endosperm color in maize. **Crop Science**, v. 29, n. 6, p. 1478–1480, 1989. DOI: 10.2135/cropsci1989.0011183X002900060031x.

SCOTT, G. E.; NELSON, L. R. Effectiveness of Resistance in Maize to Maize Dwarf Mosaic 1. **Agronomy Journal**, 1972. DOI: 10.2134/agronj1972.00021962006400030018x.

SEAL, S.; MULLER, E. Molecular analysis of a full-length sequence of a new yam badnavirus from *Dioscorea sansibarensis*. **Archives of Virology**, v. 152, n. 4, p. 819–825, 2007. DOI: 10.1007/s00705-006-0888-7.

SEAL, S. *et al.* The prevalence of badnaviruses in West African yams (*Dioscorea cayenensis-rotundata*) and evidence of endogenous pararetrovirus sequences in their genomes. **Virus Research**, v. 186, p. 144–154, 2014. DOI: 10.1016/j.virusres.2014.01.007.

SEIFERS, D. L.; MARTIN, T. J.; FELLERS, J. P. An experimental host range for Triticum mosaic virus. **Plant Disease**, v. 94, n. 9, p. 1125–1131, 2010. DOI: 10.1094/PDIS-94-9-1125.

SEO, J. K. *et al.* Complete genome sequence of longan witches' broom-associated virus, a novel member of the family Potyviridae. **Archives of Virology**, v. 162, n. 9, p. 2885–2889, 2017. a. DOI: 10.1007/s00705-017-3405-2.

SEO, J. K. *et al.* The complete genome sequence of a novel virus, bellflower veinal mottle virus, suggests the existence of a new genus within the family Potyviridae. **Archives of Virology**, v. 162, n. 8, p. 2457–2461, 2017. b. DOI: 10.1007/s00705-017-3374-5.

SHAJEELA, P. S. *et al.* Nutritional and antinutritional evaluation of wild Yam (*Dioscorea* spp.) [evaluación del valor nutricional y factores antinutricionales de *Dioscorea* spp. Silvestre]. **Tropical and Subtropical Agroecosystems**, v. 14, p. 723–730, 2011. Disponível em: <http://www.scielo.org.mx/pdf/tsa/v14n2/v14n2a30.pdf>.

SHEN, W. *et al.* The RNA-dependent RNA polymerase NiB of potyviruses plays multifunctional, contrasting roles during viral infection. **Viruses**, v. 12, n. 1, 2020. DOI: 10.3390/v12010077.

SHI, F. *et al.* Soybean Endo-1,3-Beta-Glucanase (GmGLU) Interaction With Soybean mosaic virus-Encoded P3 Protein May Contribute to the Intercellular Movement. **Frontiers in Genetics**, v. 11, n. September, p. 1–12, 2020. DOI: 10.3389/fgene.2020.536771.

SHIBOLETH, Y. M. *et al.* The Conserved FRNK Box in HC-Pro, a Plant Viral Suppressor of Gene Silencing, Is Required for Small RNA Binding and Mediates Symptom Development. **Journal of Virology**, v. 81, n. 23, p. 13135–13148, 2007. DOI: 10.1128/jvi.01031-07.

SILVA, G. *et al.* Molecular Characterization of a New Virus Species Identified in Yam (*Dioscorea* spp.) by High-Throughput Sequencing. **plants**, v. 8, n. 6, p. 167, 2019. DOI: 10.3390/plants8060167.

SIMMONDS, P. *et al.* Consensus statement: Virus taxonomy in the age of metagenomics. **Nature Reviews Microbiology**, v. 15, n. 3, p. 161–168, 2017. DOI: 10.1038/nrmicro.2016.177.

SIMÓN-BUELA, L.; GUO, H.; GARCÍA, J. Long sequences in the 5' noncoding region of plum pox virus are not necessary for viral infectivity but contribute to viral competitiveness and pathogenesis. **Virology**, v. 233, n. 1, p. 157–162, 1997. DOI: 10.1006/viro.1997.8574.

SIMPSON, J.T.; POP, M. The Theory and Practice of Genome Sequence Assembly. **Annual Review of Genomics and Human Genetics**, v. 16, p. 153–172, 2015. DOI: 10.1146/annurev-genom-090314-050032.

SIMPSON, J. T. *et al.* ABySS: A parallel assembler for short read sequence data. **Genome Research**, v. 19, n. 6, p. 1117–1123, 2009. DOI: 10.1101/gr.089532.108.

SIQUEIRA, MVB. Yam: a neglected and underutilized crop in Brazil. **Horticultura Brasileira**, v. 29, n. 1, p. 16–20, 2011. DOI: 10.1590/s0102-05362011000100003.

SOHN, Jang Il; NAM, Jin Wu. The present and future of de novo whole-genome assembly. **Briefings in Bioinformatics**, v. 19, n. 1, p. 23–40, 2018. DOI: 10.1093/bib/bbw096.

SOREL, M.; GARCIA, J. A.; GERMAN-RETANA, S. The potyviridae cylindrical inclusion helicase: A key multipartner and multifunctional protein. **Molecular Plant-Microbe Interactions**, v. 27, n. 3, p. 215–226, 2014. DOI: 10.1094/MPMI-11-13-0333-CR.

SOUZA JÚNIOR, M. T.; NICKEL, O.; GONSALVES, D. Development of virus resistant transgenic papayas expressing the coat protein gene from a Brazilian isolate of Papaya ringspot virus. **Fitopatologia Brasileira**, v. 30, n. 4, p. 357–365, 2005. DOI: 10.1590/s0100-41582005000400004.

STADLER, T. *et al.* Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). **Proceedings of the National Academy of Sciences of the United States of America**, v. 110, n. 1, p. 228–233, 2013. DOI: 10.1073/pnas.1207965110.

STAFFORD, C. A.; WALKER, G. P.; ULLMAN, D. E. Hitching a ride: Vector feeding and virus transmission. **Communicative and Integrative Biology**, v. 5, n. 1, p. 43–49, 2012. DOI: 10.4161/cib.18640.

STENGER, D. C.; FRENCH, R. Complete nucleotide sequence of Oat necrotic mottle virus: A distinct Tritimovirus species (family *Potyviridae*) most closely related to Wheat streak mosaic virus. **Archives of Virology**, v. 149, n. 3, p. 633–640, 2004. DOI: 10.1007/s00705-003-0237-z.

STENGER, D. C. *et al.* Phylogenetic relationships within the family Potyviridae: Wheat streak mosaic virus and brome streak mosaic virus are not members of the genus Rymovirus. **Phytopathology**, v. 88, n. 8, p. 782–787, 1998. DOI: 10.1094/PHYTO.1998.88.8.782.

STEWART, L. R. *et al.* Viruses in Maize and Johnsongrass in Southern Ohio. **Phytopathology**, v. 104, n. 12, p. 1360–1369, 2014. DOI: 10.1094/PHYTO-08-13-0221-R.

SUEHIRO, N. *et al.* An important determinant of the ability of Turnip mosaic virus to infect Brassica spp. and/or Raphanus sativus is in its P3 protein. **Journal of General Virology**, v. 85, n. 7, p. 2087–2098, 2004. DOI: 10.1099/vir.0.79825-0.

SUKAL, A. *et al.* 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**, v. 238, n. June, p. 29–34, 2017. DOI: 10.1016/j.virusres.2017.05.027.

SUN, P. *et al.* Structural determinants of tobacco vein mottling virus protease substrate specificity. **Protein Science**, v. 19, n. 11, p. 2240–2251, 2010. DOI: 10.1002/pro.506.

SUSAIMUTHU, J. *et al.* A member of a new genus in the Potyviridae infects *Rubus*. **Virus Research**, v. 131, n. 2, p. 145–151, 2008. DOI: 10.1016/j.virusres.2007.09.001.

TANNER, M. A.; WONG, W. H. The calculation of posterior distributions by data augmentation. **Journal of the American Statistical Association**, v. 82, n. 398, p. 528–540, 1987. DOI: 10.1080/01621459.1987.10478458.

TATINENI, S. *et al.* Triticum mosaic poacevirus enlists P1 rather than HC-Pro to suppress RNA silencing-mediated host defense. **Virology**, v. 433, n. 1, p. 104–115, 2012. DOI: 10.1016/j.virol.2012.07.016.

THORNE, J. L.; KISHINO, H.; PAINTER, I. S. Estimating the rate of evolution of the rate of molecular evolution. **Molecular Biology and Evolution**, v. 15, n. 12, p. 1647–1657, 1998. DOI: 10.1093/oxfordjournals.molbev.a025892.

THOUVENEL, J. C.; FAUQUET, C. Yam mosaic, a new potyvirus infecting *Dioscorea cayenensis* in the Ivory Coast. **Ann. appl. Biol.**, v. 93, p. 279–283, 1979.

TOLER, R. W. Maize Dwarf Mosaic, the Most Important Virus Disease of Sorghum. **Plant Disease**, v. 69, n. 11, p. 1011, 1985. DOI: 10.1094/pd-69-1011.

TOMOKO, O. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. **Journal of Molecular Evolution**, v. 40, n. 1, p. 56–63, 1995. DOI: 10.1007/BF00166595.

TORRANCE, L. *et al.* An unusual structure at one end of potato potyvirus particles. **Journal of Molecular Biology**, v. 357, n. 1, p. 1–8, 2006. DOI: 10.1016/j.jmb.2005.12.021.

TSAI, I. J.; OTTO, T. D.; BERRIMAN, M. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. **Genome Biology**, v. 11, n. 4, 2010. DOI: 10.1186/gb-2010-11-4-r41.

UMBER, M. *et al.* Molecular viral diagnosis and sanitation of yam genetic resources: Implications for safe yam germplasm exchange. **Viruses**, v. 12, n. 10, 2020. DOI: 10.3390/v12101101.

UMBER, M. *et al.* 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**, v. 162, n. 2, p. 517–521, 2017. DOI: 10.1007/s00705-016-3113-3.

VALLI, A. *et al.* A Novel Role of the Potyviral Helper Component Proteinase Contributes To Enhance the Yield of Viral Particles. **Journal of Virology**, v. 88, n. 17, p. 9808–9818, 2014. DOI: 10.1128/jvi.01010-14.

VALLI, A.; GARCÍA, J. A.; LÓPEZ-MOYA, J. J. Potyviridae. **eLS**, p. 1–10, 2015. DOI: 10.1002/9780470015902.a0000755.pub3.

VALLI, A. A. *et al.* The HCPro from the Potyviridae family: an enviable multitasking Helper Component that every virus would like to have. **Molecular Plant Pathology**, v. 19, n. 3, p. 744–763, 2018. DOI: 10.1111/mpp.12553.

VALOUEV, A. *et al.* A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. [*S. l.*], v. 18, n. 17, p. 1051–63, 2008. DOI: 10.1101/gr.076463.108.

VAUGHAN, T. G.; DRUMMOND, A. J. A stochastic simulator of birth-death master equations with application to phylodynamics. **Molecular Biology and Evolution**, v. 30, n. 6, p. 1480–1493, 2013. DOI: 10.1093/molbev/mst057.

VAUGHAN, T. G. *et al.* Efficient Bayesian inference under the structured coalescent. **Bioinformatics**, v. 30, n. 16, p. 2272–2279, 2014. DOI: 10.1093/bioinformatics/btu201.

VENTER, J. C. *et al.* The Sequence of the Human Genome. **Science**, v. 291, n. February, p. 1–49, 2001. DOI: 10.1126/science.1058040.

VILLAMOR, D. E. V. *et al.* High throughput sequencing for plant virus detection and discovery. **Phytopathology**, v. 109, n. 5, p. 716–725, 2019. DOI: 10.1094/PHYTO-07-18-0257-RVW.

WALTERMANN, A.; MAISS, E. Detection of 6K1 as a mature protein of 6 kDa in plum pox virus-infected *Nicotiana benthamiana*. **Journal of General Virology**, v. 87, n. 8, p. 2381–2386, 2006. DOI: 10.1099/vir.0.81873-0.

WANG, A. Virus and Host Plant Interactions. **eLS**, p. 1–9, 2018. DOI: 10.1002/9780470015902.a0000758.pub3.

WANG, Q. C.; VALKONEN, J. P. T. Elimination of two viruses which interact synergistically from sweetpotato by shoot tip culture and cryotherapy. **Journal of Virological Methods**, v. 154, n. 1–2, p. 135–145, 2008. DOI: 10.1016/j.jviromet.2008.08.006.

WANG, Q. C.; VALKONEN, J. P. T. Cryotherapy of shoot tips: novel pathogen eradication method. **Trends in Plant Sci**, v. 14, n. 3, p. 119–122, 2009. DOI: 10.1016/j.tplants.2008.11.010.

WANG, Q. *et al.* Cryotherapy of potato shoot tips for efficient elimination of Potato leafroll virus (PLRV) and Potato virus Y (PVY). **Potato Research**, v. 49, n. 2, p. 119–129, 2006. DOI: 10.1007/s11540-006-9011-4.

WANG, Q. *et al.* Elimination of grapevine virus A (GVA) by cryopreservation of in vitro-grown shoot tips of *Vitis vinifera* L. **Plant Science**, v. 165, n. 2, p. 321–327, 2003. DOI: 10.1016/S0168-9452(03)00091-8.

WCVP. **World Checklist of Vascular Plants, version 2.0**. Facilitated by the Royal Botanic Gardens, Kew. Disponível em: <<http://wcvp.science.kew.org/>>. Acesso em: 14 Jan. 2022.

WEI, T.; WANG, A. Biogenesis of Cytoplasmic Membranous Vesicles for Plant Potyvirus Replication Occurs at Endoplasmic Reticulum Exit Sites in a COPI- and COPII-Dependent Manner. **Journal of Virology**, v. 82, n. 24, p. 12252–12264, 2008. DOI: 10.1128/jvi.01329-08.

WEI, T. *et al.* Formation of complexes at plasmodesmata for potyvirus intercellular movement is mediated by the viral protein P3N-PIPO. **PLoS Pathogens**, v. 6, n. 6, 2010. DOI: 10.1371/journal.ppat.1000962.

WEN, R. H.; MAROOF, M. A. S.; HAJIMORAD, M. R. Amino acid changes in P3, and not the overlapping pipo-encoded protein, determine virulence of Soybean mosaic virus on functionally immune Rsv1-genotype soybean. **Molecular Plant Pathology**, v. 12, n. 8, p. 799–807, 2011. DOI: 10.1111/j.1364-3703.2011.00714.x.

WESTWOOD, J. H. *et al.* Interference with jasmonic acid-regulated gene expression is a general property of viral suppressors of RNA silencing but only partly explains virus-induced changes in plant-aphid interactions. **Journal of General Virology**, v. 95, n. PART3, p. 733–739, 2014. DOI: 10.1099/vir.0.060624-0.

WHITE, K. A. The polymerase slips and PIPO exists. **EMBO reports**, v. 16, n. 8, p. 885–886, 2015. DOI: 10.15252/embr.201540871.

WINTER, S. *et al.* Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. **Journal of General Virology**, v. 91, n. 5, p. 1365–1372, 2010. DOI: 10.1099/vir.0.014688-0.

WREN, J. D. *et al.* Plant virus biodiversity and ecology. **PLoS Biology**, v. 4, n. 3, p. 0314–0315, 2006. DOI: 10.1371/journal.pbio.0040080.

WU, C. H.; SUCHARD, M. A.; DRUMMOND, A. J. Bayesian selection of nucleotide substitution models and their site assignments. **Molecular Biology and Evolution**, v. 30, n. 3, p. 669–688, 2013. DOI: 10.1093/molbev/mss258.

WYLIE, S. J. *et al.* ICTV virus taxonomy profile: Potyviridae. **Journal of General Virology**, v. 98, n. 3, p. 352–354, 2017. DOI: 10.1099/jgv.0.000740.

WYLIE, S. J. *et al.* Caladenia virus A, an unusual new member of the family Potyviridae from terrestrial orchids in Western Australia. **Archives of Virology**, v. 157, n. 12, p. 2447–2452, 2012. DOI: 10.1007/s00705-012-1452-2.

XIE, W. *et al.* Improving marginal likelihood estimation for bayesian phylogenetic model selection. **Systematic Biology**, v. 60, n. 2, p. 150–160, 2011. DOI: 10.1093/sysbio/syq085.

XU, D. L. *et al.* Complete nucleotide sequence and taxonomy of Sugarcane streak mosaic virus, member of a novel genus in the family Potyviridae. **Virus Genes**, v. 40, n. 3, p. 432–439, 2010. DOI: 10.1007/s11262-010-0457-8.

YANG, K. *et al.* Analysis of the complete genomic sequence of a novel virus, areca palm necrotic spindle-spot virus, reveals the existence of a new genus in the family Potyviridae. **Archives of Virology**, v. 163, n. 12, p. 3471–3475, 2018. DOI: 10.1007/s00705-018-3980-x.

YANG, K. *et al.* Areca palm necrotic ringspot virus, classified within a recently proposed genus arepavirus of the family Potyviridae, is associated with necrotic ringspot disease in areca palm. **Phytopathology**, v. 109, n. 5, p. 887–894, 2019. DOI: 10.1094/PHYTO-06-18-0200-R.

YANG, Z. Among-site rate variation and its impact on phylogenetic analyses. **Trends in Ecology and Evolution**, v. 11, n. 9, p. 367–372, 1996. DOI: 10.1016/0169-5347(96)10041-0.

YANG, Z.; RANNALA, B. Molecular phylogenetics: Principles and practice. **Nature Reviews Genetics**, v. 13, n. 5, p. 303–314, 2012. DOI: 10.1038/nrg3186.

YAP, V. B.; SPEED, T. Rooting a phylogenetic tree with nonreversible substitution models. **BMC Evolutionary Biology**, v. 5, p. 1–8, 2005. DOI: 10.1186/1471-2148-5-2.

YEH, S. D. Evaluation of Induced Mutants of Papaya Ringspot Virus for Control by Cross Protection. **Phytopathology**, v. 74, n. 9, p. 1086, 1984. DOI: 10.1094/phyto-74-1086.

YODER, A. D.; YANG, Z. Estimation of primate speciation dates using local molecular clocks. **Molecular Biology and Evolution**, v. 17, n. 7, p. 1081–1090, 2000. DOI: 10.1093/oxfordjournals.molbev.a026389.

YOSHIDA, N. *et al.* Variability in the P1 gene helps to refine phylogenetic relationships among leek yellow stripe virus isolates from garlic. **Archives of Virology**, v. 157, n. 1, p. 147–153, 2012. DOI: 10.1007/s00705-011-1132-7.

YOU, B. J. *et al.* Engineered mild strains of Papaya ringspot virus for broader cross protection in cucurbits. **Phytopathology**, v. 95, n. 5, p. 533–540, 2005. DOI: 10.1094/PHYTO-95-0533.

YOU, Y.; SHIRAKO, Y. Bymovirus reverse genetics: Requirements for RNA2-encoded proteins in systemic infection. **Molecular Plant Pathology**, v. 11, n. 3, p. 383–394, 2010. DOI: 10.1111/j.1364-3703.2010.00613.x.

YUAN, W. *et al.* Complete genomic sequence of common reed chlorotic stripe virus, a novel member of the family Potyviridae. **Archives of Virology**, v. 162, n. 11, p. 3541–3544, 2017. DOI: 10.1007/s00705-017-3454-6.

YUTIN, N. *et al.* Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. **Nature Microbiology**, v. 3, n. 1, p. 38–46, 2018. DOI: 10.1038/s41564-017-0053-y.

ZAPPI, D. C. *et al.* Growing knowledge: An overview of Seed Plant diversity in Brazil. **Rodriguesia**, v. 66, n. 4, p. 1085–1113, 2015. DOI: 10.1590/2175-7860201566411.

ZECHMANN, B.; MÜLLER, M.; ZELLNIG, G. Cytological modifications in zucchini yellow mosaic virus (ZYMV)-infected Styrian pumpkin plants. **Archives of Virology**, v. 148, n. 6, p. 1119–1133, 2003. DOI: 10.1007/s00705-003-0005-0.

ZERBINO, D. R.; BIRNEY, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. **Genome Research**, v. 18, n. 5, p. 821–829, 2008. DOI: 10.1101/gr.074492.107.

ZHANG, C. *et al.* Cytoplasmic inclusion cistron of Soybean mosaic virus serves as a virulence determinant on Rsv3-genotype soybean and a symptom determinant. **Virology**, v. 391, n. 2, p. 240–248, 2009. DOI: 10.1016/j.virol.2009.06.020.

ZHANG, J.; ROBERTS, R.; RAKOTONDRAFARA, A. M. The role of the 5' untranslated regions of Potyviridae in translation. **Virus Research**, v. 206, p. 74–81, 2015. DOI: 10.1016/j.virusres.2015.02.005.

ZHANG, Y. Z. *et al.* Expanding the RNA Viroisphere by Unbiased Metagenomics. **Annual Review of Virology**, v. 6, p. 119–139, 2019. DOI: 10.1146/annurev-virology-092818-015851.

ZOU, Cheng Wu; MENG, Jiao Rong; YAO, Zi Ting; ZHANG, Lei; WANG, Zhi Qiang; WEI, Ben Hui; CHEN, Bao Shan. Genetic diversity and genome recombination in Yam mild mosaic virus isolates. **Phytopathology Research**, v. 2, n. 1, 2020. DOI: 10.1186/s42483-020-00051-0.

ZUCKERKANDL, E.; PAULING, L. **Evolutionary Divergence and Convergence, in Proteins**. Montpellier. DOI: 10.1209/epl/i1998-00224-x.

ZUCKERKANDL, E.; PAULING, L. Evolutionary Divergence and Convergence in Proteins. **Evolving Genes and Proteins**, p. 97–166, 1965. b. DOI: 10.1016/b978-1-4832-2734-4.50017-6.

CAPÍTULO II

Characterization of yam mosaic viruses from Brazil reveals a new phylogenetic group and possible incursion from the African continent

Publication: **Virus Genes**

Type of Paper: Original Article

Submission date: 09 dec 2021

Impact factor: 2.33 (2020)

**Characterization of yam mosaic viruses from Brazil reveals a new phylogenetic group
and possible incursion from the African continent**

Alejandro Risco Mendoza^{1,2}, Paolo Margaria³, Tatsuya Nagata⁴, Stephan Winter³, Rosana Blawid¹

¹ Department of Agronomy, Fitossanidade, Laboratory of Phytovirology, Federal Rural University of Pernambuco, Brazil. alejandro.mendoza@ufrpe.br, rosana.blawid@ufrpe.br

² Department of Plant Pathology, Agronomy faculty, Universidad Nacional Agraria La Molina, Perú, alejandrорisco@lamolina.edu.pe

³ Plant Virus Department, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH, Germany. stephan.winter@dsmz.de, paolo.margaria@dsmz.de

⁴ Department of Cell Biology, Laboratory of Electron Microscopy and Virology, University of Brasília, Distrito Federal, Brazil. tatsuya@unb.br

Abstract

Yam (*Dioscorea* spp.) is an important crop for smallholder farmers in the Northeast region of Brazil. Wherever yam is grown, diseases caused by yam mosaic virus (YMV) are prevalent. In the present study the diversity of YMV infecting *Dioscorea cayennensis-rotundata* was analyzed. In addition, five species of *Dioscorea* (*D. alata*, *D. altissima*, *D. bulbifera*, *D. subhastata* and *D. trifida*) commonly found in Brazil were analysed using high-throughput sequencing (HTS). Four YMV sequences were assembled from HTS as well Sanger data and phylogenetic analysis grouped them in a distinct clade. Temporal phylogenetic analyses estimated the mean evolutionary rate for the CP gene of YMV as 1.76×10^{-3} substitutions per site per year, and the time to the most recent common ancestor as 168.68 years (95% HPD: 48.56-363.28), with a most likely geographic origin in the African continent. The data presented in this study contribute to reveal key aspects of the probable epidemiological history of YMV in Brazil.

Keywords: *Yam mosaic virus*, *Dioscorea*, HTS, Phylogeographic and Geospatial studies

1. Introduction

Yam (*Dioscorea* L., family Dioscoreaceae) comprises around 629 species that grow in tropical and temperate regions, and about 140 species are found in Brazil [1] of which *D. alata*, *D. bulbifera* and *D. cayennensis-rotundata* are widely cultivated [1–4]. Yam is vegetatively propagated through tubers and this stipulates the accumulation of viruses from diverse genera: *Potyvirus*, *Badnavirus*, *Cucumovirus*, *Carlavirus*, *Comovirus*, *Fabavirus*, *Macluravirus*, *Aureusvirus* and *Potexvirus* [5–10]. Among all, yam mosaic virus (YMV, genus *Potyvirus*) is the most prevalent virus in *D. cayennensis-rotundata* [11]. YMV infection is associated with several symptoms, including mosaic, mottling, green vein banding, leaf deformation, and stunted growth leading to reduced tuber yield [7]. In Brazil, yam mosaic virus, yam mild mosaic virus (YMMV) and dioscorea bacilliform viruses (DBVs, genus *Badnavirus*) are widespread in yam producing areas [14, 15] and all three viruses can occur either in single or mixed infections to cause disease [15].

YMV is a member of the genus *Potyvirus* with a positive single-stranded RNA genome [16], approximately 9.6 kb in length [7, 17]. Its genome organization is well known from the first complete genome assembly of an isolate from Ivory Coast [17] and from an isolate from Nigeria [7], which are the only two complete genome sequences determined so far.

The molecular variability of P1, HC-Pro, P3, NIb gene sequences, the complete CP and its N-terminus (N-Ter) [18], the C-terminus (C-Ter) of NIb and the 3'-UTR [7, 18–20], has been assessed to classify YMV isolates of different geographical origin into phylogroups. So far, nine different groups (I–IX) have been described: group I includes isolates from Burkina Faso; group II from Guadeloupe and Ivory Coast; group III from Benin and Cameroon [18]; group IV from Benin, Burkina Faso, and Cameroon; group V from French Guiana; group VI from Guadeloupe and Puerto Rico; group VII from Nigeria; group VIII from Guadeloupe, and group IX from Cameroon [20].

In Brazil, the presence of YMV in northeastern Brazil was reported since 2006 through serological tests, and recently a survey of the Embrapa germplasm collection showed a high incidence of YMV-infected plants [21]. Because there was no further genome data available, in this study we assembled four YMV genomic sequences using high-throughput sequencing (HTS) and sanger sequencing data obtained from YMV-infected *Dioscorea* species cultivated in the Northeast and Southeast regions of Brazil. Evolutionary analyses were conducted using various genomic regions, to describe the molecular variability among the assembled sequences. In a Bayesian phylodynamic study based on CP gene sequences, evolutionary rate and

timeframe to the most recent common ancestor (TMRCA) were calculated, to contribute to clarify the likely epidemiological history of YMV in Brazil. .

2. Material and Methods

2.1. Field surveys

From October 2018 to January 2020 surveys were carried out to reveal the presence of YMV and YMMV in Brazilian yam fields. A total of 150 symptomatic and asymptomatic young leaves (75 *D. alata*, 63 *D. cayennensis-rotundata*, and 12 *D. bulbifera*, one leaf/plant) were collected in the states of Pernambuco, Alagoas, Paraíba, Bahia, Maranhão and São Paulo, desiccated on calcium chloride (CaCl₂) and kept at 4 °C until further use. Double antibody sandwich (DAS)-ELISA was performed using the DSMZ ELISA kits (YMV, AS-1142; YMMV, AS-1038) following the laboratory routines and recommendations associated with the assays. Positive controls for YMV (PC-1142) and YMMV (PC-1214) and negative controls were included in each ELISA and samples were considered positive when absorbance values were three times the mean of absorbance values of negative controls.

2.2. High-Throughput Sequencing

A total of 35 yam samples, grouped into five pools, were sequenced by HTS: pool DR1 consisted of eight samples of *D. cayennensis-rotundata* collected exclusively in the state of Alagoas; pool DR2 of eight samples of *D. cayennensis-rotundata* from the states of Pernambuco, Paraíba and Alagoas; pool DA of seven samples of *D. alata* from the states of Pernambuco and Paraíba; pool DB of nine samples of *D. bulbifera* from Pernambuco and São Paulo; and pool ID of a single sample of each species of *D. trifida*, *D. subhastata* and *D. altissima* from the state of Pernambuco (Supplementary Table S1). A site map (Fig. 1) of the collection sites was generated using the geographical information system (GIS) software QGIS v3.10 (<https://www.qgis.org>) from the respective GPS coordinates listed in Supplementary Table S1.

Total RNA was extracted using a GenCatch™ Plant Total RNA Miniprep Kit (Epoch Life Sciences, Missouri City, USA) and quality was evaluated by agarose gel electrophoresis. Following ribosomal-RNA depletion using a RiboMinus™ Plant Kit (ThermoFischer Scientific, Carlsbad, USA), cDNA and second-strand synthesis was performed using random octamer primers. The libraries were constructed using a Nextera XT DNA Library Preparation Kit (Illumina Inc., San Diego, USA) and subjected to MiSeq paired-end reads sequencing (2 x

301bp). The 5'-end sequence of YMV_DR1 was confirmed by 5'-RACE according to Hayashi et al. [22].

An additional dataset generated by HiSeq sequencing (100 bp, paired-end, Illumina Inc. Macrogen, Seoul, Korea) of a mixed sample (“Dsp”) of *D. cayennensis-rotundata* and *D. alata* collected in the states of Pernambuco and Paraíba in 2014 was also considered in this work.

In order to validate assembled YMV genomic sequences from the HTS data, cDNA fragments were amplified by reverse transcription polymerase chain reaction (RT-PCR) from total RNA from a single plant of *D. cayennensis-rotundata* from Alagoas state and sequence by sanger. The 5' and 3'-terminal ends were amplified following standard protocols for rapid amplification of cDNA ends (RACE). Primers used to amplify the whole genomic sequence of YMV are indicated in Supplementary Table S2.

Supplementary

Table S1. Samples of *Dioscorea* species collected from different locations in Brazil.

Pool	Species	Collection site	latitude	longitude	YMV	YMMV
DR1	<i>D. cayennensis-rotundata</i>	Chã preta, Alagoas	9°14'18.0"S	36°17'32.2"W	+	-
	<i>D. cayennensis-rotundata</i>	Chã preta, Alagoas	9°14'18.0"S	36°17'32.2"W	n.t.	n.t.
	<i>D. cayennensis-rotundata</i>	Mar vermelho, Alagoas	9°31'39.7"S	36°03'17.5"W	+	-
	<i>D. cayennensis-rotundata</i>	Mar vermelho, Alagoas	9°31'39.7"S	36°03'17.5"W	n.t.	n.t.
	<i>D. cayennensis-rotundata</i>	Mar vermelho, Alagoas	9°31'39.7"S	36°03'17.5"W	n.t.	n.t.
	<i>D. cayennensis-rotundata</i>	Mar vermelho, Alagoas	9°31'39.7"S	36°03'17.5"W	n.t.	n.t.
	<i>D. cayennensis-rotundata</i>	Viçosa, Alagoas	9°22'33.9"S	36°14'31.2"W	+	-
	<i>D. cayennensis-rotundata</i>	Viçosa, Alagoas	9°22'33.9"S	36°14'31.2"W	n.t.	n.t.
DR2	<i>D. cayennensis-rotundata</i>	Alhandra, Paraíba	7°18'51.0"S	34°56'45.1"W	-	-
	<i>D. cayennensis-rotundata</i>	Bonito, Pernambuco	8°32'10.7"S	35°41'59.5"W	+	-
	<i>D. cayennensis-rotundata</i>	Bonito, Pernambuco	8°32'10.7"S	35°41'59.5"W	n.t.	n.t.
	<i>D. cayennensis-rotundata</i>	Camocim, Pernambuco	8°19'57.8"S	35°46'16.4"W	n.t.	n.t.
	<i>D. cayennensis-rotundata</i>	Chã preta, Alagoas	9°14'18.0"S	36°17'32.2"W	n.t.	n.t.
	<i>D. cayennensis-rotundata</i>	Chã preta, Alagoas	9°14'18.0"S	36°17'32.2"W	n.t.	n.t.
	<i>D. cayennensis-rotundata</i>	Chã preta, Alagoas	9°14'18.0"S	36°17'32.2"W	n.t.	n.t.
	<i>D. cayennensis-rotundata</i>	Mar vermelho, Alagoas	9°31'39.7"S	36°03'17.5"W	+	-
DA	<i>D. alata</i>	Bonito, Pernambuco	8°32'10.7"S	35°41'59.5"W	-	+
	<i>D. alata</i>	Bonito, Pernambuco	8°48'20.2"S	36°00'48.4"W	-	+
	<i>D. alata</i>	Bonito, Pernambuco	8°32'10.7"S	35°41'59.5"W	-	-
	<i>D. alata</i>	Quipapa, Pernambuco	8°48'20.2"S	36°00'48.4"W	-	+
	<i>D. alata</i>	Quipapa, Pernambuco	8°48'20.2"S	36°00'48.4"W	-	+
	<i>D. alata</i>	Quipapa, Pernambuco	8°48'20.2"S	36°00'48.4"W	-	+
	<i>D. alata</i>	Marcaças, Paraíba	6°46'02.1"S	35°00'53.6"W	-	+
	<i>D. bulbifera</i>	Aldeia, Pernambuco	7°58'09.8"S	34°59'31.9"W	-	-
	<i>D. bulbifera</i>	Aldeia, Pernambuco	7°56'10.5"S	35°01'04.4"W	-	-

DB	<i>D. bulbifera</i>	Aldeia, Pernambuco	7°56'10.5"S	35°01'04.4"W	-	-
	<i>D. bulbifera</i>	Aldeia, Pernambuco	7°58'09.8"S	34°59'31.9"W	-	-
	<i>D. bulbifera</i>	Barra de guabiraba, Pernambuco	8°25'10.6"S	35°39'42.4"W	-	-
	<i>D. bulbifera</i>	Bonito, Pernambuco	8°32'10.7"S	35°41'59.5"W	-	-
	<i>D. bulbifera</i>	Brennand Forest, Pernambuco	8°02'24.9"S	34°58'54.0"W	-	-
	<i>D. bulbifera</i>	Rio Claro, São Paulo	22°23'32.6"S	47°33'45.7"W	-	-
	<i>D. bulbifera</i>	Rio Claro, São Paulo	7°56'10.5"S	35°01'04.4"W	-	-
ID	<i>D. altissima</i>	Brennand Forest, Pernambuco	8°02'24.9"S	34°58'54.0"W	n.t.	n.t.
	<i>D. subhastata</i>	Brennand Forest, Pernambuco	8°02'24.9"S	34°58'54.0"W	n.t.	n.t.
	<i>D. trifida</i>	Brennand Forest, Pernambuco	8°02'24.9"S	34°58'54.0"W	n.t.	n.t.

Positive (+) and negative (-) ELISA-test for YMV and YMMV. Not tested (n.t.).

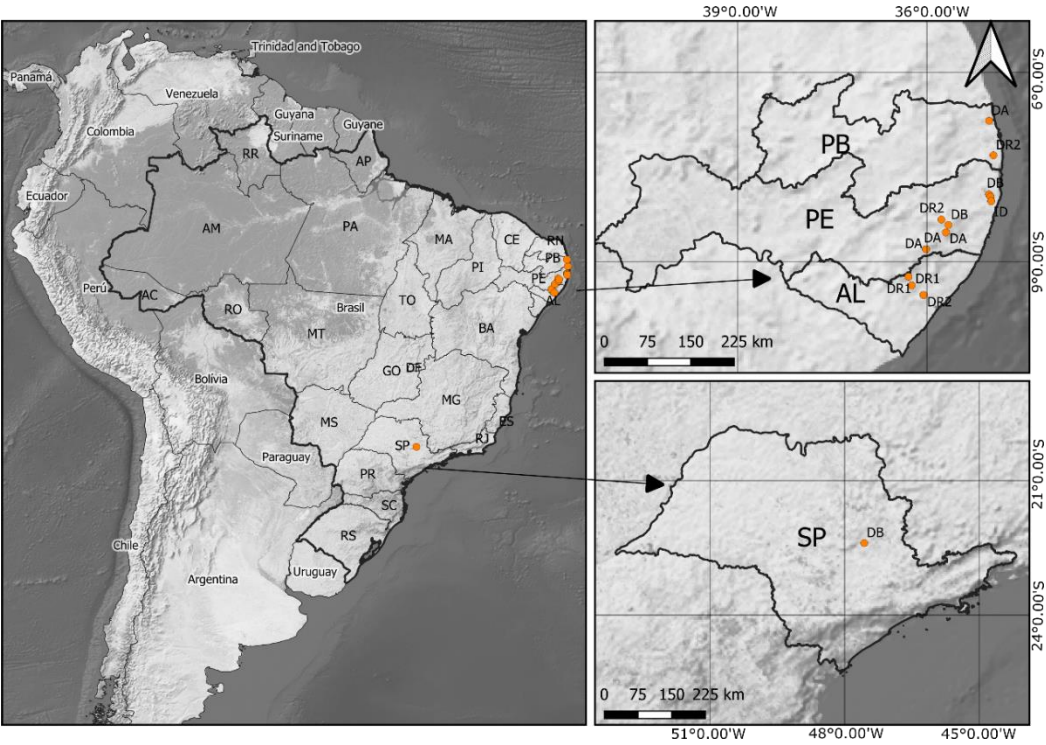


Fig. 1 Map showing the collection sites (orange dots) of *Dioscorea* species from different states of Brazil. DR: *D. cayennensis-rotundata*, DA: *D. alata*, DB: *D. bulbifera*, ID: *D. trifida*, *D. subhastata*, and *D. altissima*.

Supplementary

Table S2. Primers used to amplify whole genomic sequence of YMV

A. amplicons sequenced by sanger.

Amplicon	Primer name	Nucleotide position ^a		Amplicon size (bp)
For RACE 5'	YMV663-R	1	498	498
Amplicon A	YMV955-F / YMV4026-R	429	3428	3000
Amplicon B	YMV3732-F / YMV6136-R	3139	5511	2373
Amplicon C	YMV5925-F / YMV8371-R	5301	7696	2396
Amplicon D	YMV8205-F / oligo-d(T)50 M4	7537	9608	2072

B. Primers used to amplify genomic sequences

Name	Sequence (5'-3')	Nucleotide position	
YMV663-R	TGAATGTGAGTGATAAGAGATGG	476	498
YMV955-F	AGAAGGCCAAAACAGGAGAGAGAG	429	451
YMV2180-F	GCTTACATGGAGACCACACAAAC	1205	1227
YMV3732-F	TGCAATTCGATATGTGTTTCCTG	3139	3161
YMV3955-R	CTCTTTATGCGCGTCAAGAAGTT	2765	2787
YMV4026-R	CTTCATCAGCACACGCATGTTTA	3406	3428
YMV5332-F	TGAGAGGGTTTAATCGTTTCGGT	4002	4024
YMV5925-F	ATGTGTGTTTCGAGTTCTATGTTC	5301	5323
YMV6136-R	TTCTCTGATTGATGATGAAGTGC	5489	5511
YMV7730-F	AGGGTGAATTGAGACAAACTGGA	6147	6169
YMV8205-F	AACTAGAACATTCACAGCAGCAC	7537	7559
YMV8371-R	CAATTTCGCATCAATTCATTCC	7674	7696
YMV8892-F	AATTAGGCCAGATATGGAGCACC	8053	8075
YMV9648-F	GAATGTGGACGATGATGGACGG	8728	8750
YMV9877-R	TGAGTTCCGCAATTCATGTACG	9147	9169
YMV9989-F	GAATGTGGACGATGATGGACGG	9030	9051

^a Position of PCR primers on the genomic sequence of YMV_DR1

2.3. Virome analysis

The sequence analysis pipeline described by Blawid et al. [23] was employed for a global virome analysis. In brief, reads were quality-checked with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and trimmed with Trimmomatic v.0.39 [24] implemented in the Galaxy platform Version 0.38.0 (<https://usegalaxy.org/>), according to the following parameters: ILLUMINACLIP:Nextera:2:30:10, LEADING:3, TRAILING:3, SLIDINGWINDOW: 4:20, MINLEN:36. Contigs were assembled with SPAdes v.3.12.0 [25] using *K-mer* values of 21,33,55,77 and subjected to tBlastX searches (E-value $1e^{-10}$, Matrix BLOSUM62, Word Size 3) against the viral reference genome database (RefSeq

March 2020, NCBI) implemented in Geneious v. R11 (Biomatters, Auckland, New Zealand). Sequences resulting in a hit in tBlastX with identities equal or higher than 70% to viral sequences were selected and *de novo* assembled in Geneious using “map-to-reference” function (parameters: medium-low/fast sensitivity, fine-tuning of iteration up to five times). The Sequence Demarcation Tool (SDTv.1.2) [26] was used for pairwise comparison analyses at the nucleotide and amino acid levels with other sequences retrieved from GenBank.

2.4. Bayesian inferences

Two datasets, “11” and “32”, generated by retrieving YMV sequences available from GenBank and those assembled in this work, were used for a Bayesian analysis. For dataset 11, a partition of four independent dataset analyses was carried out in order to compare different regions of the YMV genome: we considered 11 sequences for P1 (“dataset_11_P1”, 723 aligned sites), HC_Pro (“dataset_11_HC”, 459 sites), P3 (“dataset_11_P3”, 390 aligned sites) and N1b (“dataset_11_N1b”, 927 sites) (Supplementary Table S3).

The “dataset_32_CP” (909 aligned sites) consisted of 28 sequences retrieved from GenBank and the four YMV sequences from Brazil assembled in this work. This dataset was employed for the Bayesian and the geospatial analyses (Supplementary Table S3). To compare and verify the structure of the phylogenetic groups generated by the “dataset_32_CP”, three additional datasets were also employed in this study: (i) dataset_32_CP-NTer with 450 aligned sites containing the N-terminus region of CP; (ii) dataset_32_CP-CTer containing the CPcore/C-terminus (720 aligned sites), and (iii) dataset_32_N1b-CP-UTR including the partial N1b, the complete CP and 3'-UTR regions with 1183 aligned sites.

Each dataset sequences were aligned using the MAFFT v.7 algorithm [27] and the statistical selection of best-fit models of nucleotide substitutions was assessed with jModelTest v.2.1.10 [28]. The General Time-Reversible (GTR) model was selected with a fraction of invariant (I) sites and a gamma distribution (G) of substitution rates. Phylogenetic trees were constructed using MrBayes v.3.2.7 [29], which implements a Markov chain Monte Carlo (MCMC) approach to approximate posterior probabilities of phylogenetic trees and clades. Each run was based on a chain of 2×10^6 generations, with a tree sampled every 100 generations (mcmc ngen = 2,000,000 samplefreq = 100). Plots of tree topology likelihoods versus generation numbers were examined with Tracer v.1.7.1 [30] and the first 10 % of samples were determined as burn-in. Runs were ceased when the standard deviation of split frequencies was below 0.01 and the Potential Scale Reduction Factor (PSRF) for the posterior probability distribution reached values of 1.0 [29]. Phylogenetic trees were visualized and edited with

160 FigTree v.1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>) and Inkscape v. 1.0.2-2
161 (<https://www.inkscape.org/>).

162 **Supplementary**163 **Table S3.** YMV sequences used for the Bayesian inferences. In bold, sequences from Brazil.

Dataset	Region (Sequences)	Nucleotides (aligned sites)	Amino acid (position)	GenBank Accession numbers/ sequences from Brazil ^a
dataset_11_P1	P1	723 nt	FGQ ₈ F - [H/G]LT ₂₄₅ [H/Y]	YMU53426, YMU53422, YMU53423, YMU53424, YMU53425, MG711313, NC_004752, YMV_DR1^a , YMV_DR2^a , YMV_Dsp^a , YMV_I4^a
dataset_11_HC	HC-Pro	459 nt	MWG ₅₀₇ Q - DEK ₆₅₆ D	YMU53427, YMU53428, YMU53429, YMU53430, YMU53431, MG711313, NC_004752, YMV_DR1^a , YMV_DR2^a , YMV_Dsp^a , YMV_I4^a
dataset_11_P3	P3	390 nt	AFI ₈₂₈ S - TAD ₉₅₄ F	YMU53432, YMU53433, YMU53434, YMU53435, YMU53436, MG711313, NC_004752, YMV_DR1^a , YMV_DR2^a , YMV_Dsp^a , YMV_I4^a
dataset_11_NIb	NIb	927 nt	[Q/L]FY ₂₄₉₅ S - VSH ₂₈₀₀ Q	YMU53437, YMU53438, YMU53439, YMU5340, YMU5341, MG711313, NC_004752, YMV_DR1^a , YMV_DR2^a , YMV_Dsp^a , YMV_I4^a
dataset_32_CP	CP	909 nt	ADT ₂₈₀₄ Q - MRG ₃₁₀₂ M	MG711313, NC_004752, AJ238401, AJ244042, AJ244043, AJ244044, AJ244045, AJ244046, AJ244047, AJ244048, AJ244049, AJ244050, AJ244051, AJ244052, AJ244053, AJ244054, AJ244055, AJ244056, AJ244057, AJ244058, AJ244059, AJ244060, AJ244061, AJ244062, AJ244063, AJ244064, AJ244065, AJ244066, YMV_DR1^a , YMV_DR2^a , YMV_Dsp^a , YMV_I4^a
dataset_32_CP-NTer	CP: N-Ter + Partial Core	450 nt	ADT ₂₈₀₄ Q - MNG ₂₉₅₀ L	
dataset_32_CP-CTer	CP: Partial Core + C-Ter	720 nt	T[H/R]D ₂₈₆₆ [R/K] - MRG ₃₁₀₂ M	
dataset_32_NIb-CP-UTR	Partial NIb + CP + 3'-UTR	1183 nt	From ALY ₂₇₆₈ T	

^a Accession number of the sequences from Brazil: YMV_DR1 (OK239700), YMV_DR2 (OK239701), YMV_Dsp (OK239702), and YMV_I4 (OL739290)

164

2.5. Geospatial analysis

The geospatial analysis was performed on “dataset_32_CP”. The integration of digital world map file “shp” format, the Bayesian phylogenetic tree, georeferenced locations, host, country, and phylogroup data were merged into a single interactive visualization and supported by statistical tools included in GenGIS v.2.5 [31].

2.6. Evolutionary and temporal analysis

A selection of 26 CP sequences, among the ones present in “dataset_32_CP”, that showed the highest correlation for the temporal signal was used for the evolutionary analysis (“dataset_26”, Supplementary Table S4). Evaluation of temporal molecular evolutionary signals through root-to-tip linear regression analyses between genetic divergence and sampling date using the best-fitting root were conducted using TempEst v.1.5.3 [32]. Next, phylogenetic signals from aligned sequences were assessed by testing nucleotide saturation substitutions using DAMBE v.6 [33]. Maximum-likelihood (ML) phylogenetic trees were generated using MEGAX [34] under a GTR+I+G nucleotide substitution model. Branch supports were inferred using 1000 bootstrap replicates. To assess recombination events from the “dataset_26” sequences, we employed the default recombination detection methods using the Recombination Detection Program (RDP) v4.100 [35]. Bayesian evolutionary analyzes were performed with BEAST v.1.10.4 [36] implemented with the BEAGLE v.2.1.2 library [37] for improving computational efficiency. Temporal calibration was performed by the data of sampling. Also, tests of data-randomization were performed for accurate estimations of substitution rate as described by Duchêne et al. [38] (data not shown). The nucleotide substitution GTR model with estimated base frequencies and Gamma rate heterogeneity with invariant sites was employed.

We first tested the strict model, but no adequate sampling of posterior was produced. Therefore, the uncorrelated relaxed clock with a log-normal relaxed distribution (UCLN) [39] and with exponential distribution (UCEXP) [39] were tested, as well four demographic models to investigate population size changes through time: exponential growth, expansion growth, Bayesian Skyline, and Constant Size [40, 41] with growth rate parameterization (EXP) [42]. Tree priors were adjusted to typical RNA viruses (default.allMus to lognormal with initial value of 0.001, exponential coalescent growth rate to Laplace $\mu=0$; scale=100; Exponential coalescent population size to a lognormal distribution function with $\mu=1.0$; $\sigma = 5$).

Two discrete traits were studied: Host and Country under the symmetric substitution model [43]. The best fit model was chosen by marginal Likelihood comparisons using the path

sampling (PS) and stepping-stone sampling (SS) estimates [44, 45]. Analyses were performed for four independent runs of 200 million MCMC steps with sampling parameters and trees every 1,000 generations. Logcombiner software allowed log files and tree files to combine. Files were inspected with the Tracer program. The maximum clade credibility (MCC) tree was generated using Tree-Annotator v1.10.4. Tracer v.1.7.1 was employed to verify the convergence of parameters, in which an effective sample size of at least 200 was accepted [30]. FigTree v.1.4.3 was used for visualization of MCC trees and tree editing (<http://tree.bio.ed.ac.uk/software/figtree/>).

Supplementary

Table S4. Overview of the YMV CP sequences used for the evolutionary analysis (“dataset_26”).

Phylogenetic Group	Code ^a	Sequence Length (nt)	Collection date	Country	Host	Reference
I	AJ244050_BUF	873	1992	Burkina Faso	<i>D. rotundata</i>	Bousalem et al., 2000
I	AJ244051_BUF	873	1992	Burkina Faso	<i>D. rotundata</i>	Bousalem et al., 2000
I	AJ244052_BUF	873	1992	Burkina Faso	<i>D. rotundata</i>	Bousalem et al., 2000
I	AJ244053_BUF	873	1992	Burkina Faso	<i>D. rotundata</i>	Bousalem et al., 2000
II	NC_004752_CDI	909	1977	Ivory Coast	<i>D. rotundata</i>	Aleman et al. 1996
III	AJ244048_BEN	909	1994	Benin	<i>D. rotundata</i>	Bousalem et al., 2000
III	AJ244049_BEN	909	1994	Benin	<i>D. rotundata</i>	Bousalem et al., 2000
III	AJ244054_CMM	909	1994	Cameroon	<i>D. rotundata</i>	Bousalem et al., 2000
III	AJ244056_BEN	909	1994	Benin	<i>D. rotundata</i>	Bousalem et al., 2000
IV	AJ244046_BEN	909	1994	Benin	<i>D. rotundata</i>	Bousalem et al., 2000
IV	AJ244055_CMM	909	1994	Cameroon	<i>D. rotundata</i>	Bousalem et al., 2000
IV	AJ244065_BUF	909	1992	Burkina Faso	<i>D. alata</i>	Bousalem et al., 2000
IV	AJ244066_BUF	909	1992	Burkina Faso	<i>D. alata</i>	Bousalem et al., 2000
V	AJ244045_FGUI	909	1991	French Guiana	<i>D. trifida</i>	Bousalem et al., 2000
V	AJ244061_FGUI	909	1992	French Guiana	<i>D. trifida</i>	Bousalem et al., 2000
V	AJ244062_FGUI	909	1992	French Guiana	<i>D. trifida</i>	Bousalem et al., 2000
VI	AJ244043_PUR	909	1994	Puerto Rico	<i>D. alata</i>	Bousalem et al., 2000
VI	AJ244044_FGUA	909	1994	Guadeloupe	<i>D. rotundata</i>	Bousalem et al., 2000
VI	AJ244057_FGUA	909	1994	Guadeloupe	<i>D. rotundata</i>	Bousalem et al., 2000
VI	AJ244063_FGUA	909	1991	Guadeloupe	<i>D. rotundata</i>	Bousalem et al., 2000
VII	AJ244047_NIG	909	1992	Nigeria	<i>D. rotundata</i>	Bousalem et al., 2000
VII	MG711313_NIG	909	2017	Nigeria	<i>D. rotundata</i>	Bömer et al., 2019
IX	AJ244042_CMM	909	1994	Cameroon	<i>D. cayennensis-rotundata</i>	Bousalem et al., 2000
X	YMV_DR1 (OK239700)	909	2019	Brazil	<i>D. cayennensis-rotundata</i>	This work
X	YMV_DR2 (OK239701)	909	2019	Brazil	<i>D. cayennensis-rotundata</i>	This work
X	YMV_Dsp (OK239702)	909	2014	Brazil	<i>D. sp.</i>	This work

^a NCBI Accession; Country code

3. Results

3.1. Symptoms and serological detection

Symptoms of mosaic, chlorosis and leaf mottling were commonly observed in *D. cayennensis-rotundata* (Fig. 2 a and b), while *D. alata* showed only weak and barely visible mosaic symptoms on leaves (Fig. 2 c and d). All inspected *D. subhastata*, *D. altissima*, *D. trifida*, and *D. bulbifera* (Fig. 2 e - h) were free of conspicuous leaf symptoms. YMV was detected by ELISA only in *D. cayennensis-rotundata* and was present in all sampling sites (Alagoas, Bahia, Maranhão, Paraíba, and Pernambuco), while YMMV was found only in *D. alata* (Paraíba and Pernambuco). None of the *D. bulbifera* plants tested positive for any of the two viruses (Table 1).

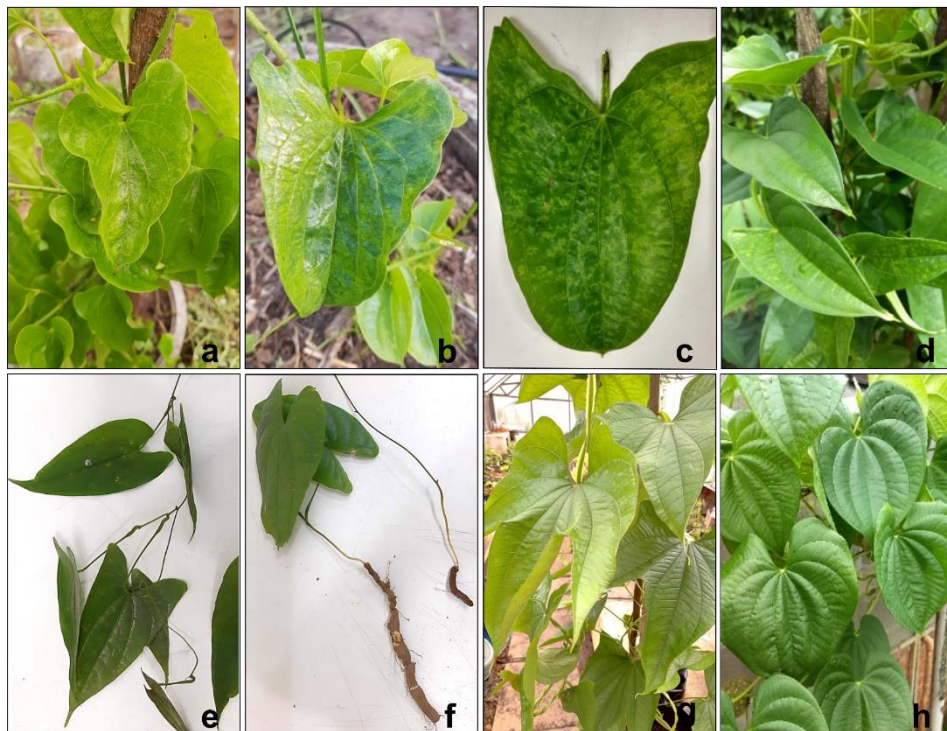


Fig. 2 *Dioscorea* species collected from yam fields in Brazil. **a** and **b**. *D. cayennensis-rotundata*. **c** and **d**. *D. alata*. **e**. *D. subhastata*. **f**. *D. altissima*. **g**. *D. trifida*. **h**. *D. bulbifera*.

Table 1 Serological tests to detect YMV and YMMV in *D. alata*, *D. cayennensis-rotundata* and *D. bulbifera* from samples collected in six different states of Brazil. Numbers inside brackets represent the number of samples that tested positive by ELISA per total number of samples, and numbers outside brackets viral incidence (%).

State/Dioscorea specie	Samples tested	YMMV		YMV	
		(+)	(%)	(+)	(%)
ALAGOAS	6	(0/6)	0	(4/6)	66.7
<i>D. cayennensis-rotundata</i>	6	(0/6)	0	(4/6)	66.7
BAHIA	3	(0/3)	0	(2/3)	66.7
<i>D. cayennensis-rotundata</i>	3	(0/3)	0	(2/3)	66.7
MARANHÃO	3	(0/3)	0	(1/3)	33.3
<i>D. alata</i>	2	(0/2)	0	(0/2)	0.0
<i>D. cayennensis-rotundata</i>	1	(0/1)	0	(1/1)	100.0
PARAÍBA	68	(25/68)	36.8	(16/30)	53.3
<i>D. alata</i>	39	(25/39)	64.1	(0/8)	0.0
<i>D. cayennensis-rotundata</i>	29	(0/29)	0	(16/22)	72.7
PERNAMBUCO	67	(28/67)	41.8	(1/28)	3.6
<i>D. alata</i>	34	(28/34)	82.4	(0/15)	0.0
<i>D. bulbifera</i>	9	(0/9)	0	(0/9)	0.0
<i>D. cayennensis-rotundata</i>	24	(0/24)	0	(1/4)	25.0
SÃO PAULO	3	(0/3)	0	(0/3)	0.0
<i>D. bulbifera</i>	3	(0/3)	0	(0/3)	0.0
Total					
<i>D. alata</i>	75	(53/75)	70.7	(0/25)	0.0
<i>D. bulbifera</i>	12	(0/12)	0	(0/12)	0.0
<i>D. cayennensis-rotundata</i>	63	(0/63)	0	(24/36)	66.7
	150	(53/150)	35.3	(24/73)	32.9

3.2. Virome analysis and characterization of YMV genomes

The list of viruses detected by HTS is presented in Table 2. YMV was assembled only from pooled *D. cayennensis-rotundata* samples (DR1 and DR2). Dioscorea mosaic-associated virus (RNA1 and RNA2) and Dioscorea bacilliform AL virus were detected in pool DR1 and DR2, while Dioscorea bacilliform RT virus was found in pool DR1. Dioscorea virus A (Family *Betaflexiviridae*) and YMMV were identified only in pool DA.

239 **Table 2** List of viral contigs assembled from each pool by HTS analysis.

Pool	ID	States	<i>Dioscorea</i> species	Genus	Species	Blastn NCBI Code	Identity (%)	Length (nt)	Coverage	Reads
DR1	YMV_DR1	AL	<i>D. cayennensis-rotundata</i>	<i>Potyvirus</i>	<i>Yam mosaic virus</i>	MG711313	90.67	9608	1463	32401
	DBALV_DR1		<i>D. cayennensis-rotundata</i>	<i>Badnavirus</i>	<i>Dioscorea bacilliform AL virus</i>	NC_038381	89.91	7546	134	1747
	DBRTV_DR1		<i>D. cayennensis-rotundata</i>	<i>Badnavirus</i>	<i>Dioscorea bacilliform RT virus</i>	MF476845	87.92	7395	101	729
	DMaV_RNA1_DR1		<i>D. cayennensis-rotundata</i>	<i>Sadwavirus</i>	<i>Dioscorea mosaic-associated virus</i> - RNA1	KU215538	89.84	5932	243	3936
	DMaV_RNA2_DR1		<i>D. cayennensis-rotundata</i>	<i>Sadwavirus</i>	<i>Dioscorea mosaic-associated virus</i> - RNA2	KU215539	90.74	3792	729	5773
DR2	YMV_DR2	AL, PB, PE	<i>D. cayennensis-rotundata</i>	<i>Potyvirus</i>	<i>Yam mosaic virus</i>	MG711313	90.55	9608	1340	31923
	DBALV_DR2		<i>D. cayennensis-rotundata</i>	<i>Badnavirus</i>	<i>Dioscorea bacilliform AL virus</i>	NC_038381	89.57	7572	138	1728
	DMaV_RNA1_DR2		<i>D. cayennensis-rotundata</i>	<i>Sadwavirus</i>	<i>Dioscorea mosaic-associated virus</i> - RNA1	KU215538	91.6	5945	491	8503
	DMaV_RNA2_DR2		<i>D. cayennensis-rotundata</i>	<i>Sadwavirus</i>	<i>Dioscorea mosaic-associated virus</i> - RNA2	KU215539	93.33	3811	470	3810
DA	YMMV_DA	PB, PE	<i>D. alata</i>	<i>Potyvirus</i>	<i>Yam mild mosaic virus</i>	JX470965	86.65	9512	358	8199
	DBALV_DA		<i>D. alata</i>	<i>Badnavirus</i>	<i>Dioscorea bacilliform AL virus</i>	KX008571	90.08	7396	48	508
	DVA_DA		<i>D. alata</i>	unclassified	<i>Dioscorea virus A</i>	LC467961	95.35	7592	510	7843
Dsp	YMV_Dsp	PE	<i>D. sp.</i>	<i>Potyvirus</i>	<i>Yam mosaic virus</i>	MG711313	90.73	9608	1565	66099

240 AL: Alagoas, PB: Paraíba, PE: Pernambuco, SP: São Paulo.

Three HTS sequences of YMV, 9,608 nt in length, were assembled from samples collected in 2020 and 2014, and were designated as YMV_DR1/YMV_DR2 and YMV_Dsp, respectively (Table 2, Supplementary Table S5). Full length genomic Sanger sequencing (YMV_I4) of 9,608 nt (without the poly(A) tail) in length was obtained from a single plant of *D. cayennensis-rotundata*. The four sequences of YMV_DR1 (OK239700), YMV_DR2 (OK239701), YMV_Dsp (OK239702) and YMV_I4 (OL739290) were deposited in the NCBI GenBank database.

The assembled YMV_I4 genomic sequence from the Sanger sequencing showed 97.9-98.4% identity with the YMV sequences assembled from the HTS data.

The polyprotein of all four YMV sequences consists of 3,103 amino acids, with an estimated molecular weight of ~350 kDa. The predicted nine cleavage sites are depicted in Figure 3. The short ORF P3N-PIPO was predicted from an adenosine-rich region between nt positions 2,863 and 2,869 (GA₂₈₆₃A_AAA_AA₂₈₆₉T), resulting in a trans-frame protein P3N-PIPO (Supplementary Table S5 and Fig. 3).

Motifs at the same protein position were identified in the four YMV sequences: the HC-Pro motifs FRNK₄₈₁ and PTK₆₀₉ involved in microRNA binding, and aphid transmission, respectively [46]; the helicase CI motifs Walker A-GxxxxGK₁₂₄₆[S/T], Walker B-hhhh[D/E₁₃₃₀] [47, 48], VATNIIENGVTL₁₄₆₉ [49], and QRLGRVGR₁₅₁₃ [50], Nlb motifs of RNA polymerases SLKAEL₂₄₅₈, WTVGMTKIFYGGWNELMRKL₂₅₂₁, VDADGSQFDSSL₂₅₃₉, KGNNSGQPSTVVDNTLMVMLAMYY₂₆₁₂, FYANGDDLII₂₆₃₉, EDLWFMSHQAKV₂₆₈₄, and YIPKLEQERIVS₂₇₀₀ [18], and the DAG₂₈₁₆ motif at the N-Terminus of CP (Supplementary Table S5).

Supplementary

Table S5. Genome characterization, putative proteolytic cleavage sites, and conserved motifs of the YMV sequences generated in this study. YMV_DR1 (OK239700), YMV_DR2 (OK239701), YMV_Dsp (OK239702), and YMV_I4 (OL739290).

Name	Type	Initial position	Final position	Length
5'-UTR	5'UTR	1	133	133
Polyprotein	CDS	134	9445	9312
P1	CDS	134	1024	891
P1/HC-Pro:IRQF ₂₉₇ /S	Cleavage site	1013	1027	15
HC-Pro	CDS	1025	2398	1374
Motif: FRNK ₄₈₁	motif	1565	1576	12
Motif: PTK ₆₀₉	motif	1952	1960	9
HC-Pro/P3: YQVG ₇₅₅ /G	trimmed	2387	2401	15
P3	CDS	2399	3448	1050
PIPO	CDS	2863	3093	231
P3/6K1: VEHQ ₁₁₀₅ /S	Cleavage site	3437	3451	15
6K1	CDS	3449	3598	150
6K1/CI: VSHQ ₁₁₅₅ /S	Cleavage site	3587	3601	15
CI	CDS	3599	5503	1905
WalkerA: GxxxxGK ₁₂₄₆ [S/T]	motif	3851	3877	27
Walker B: hhhh[DE ₁₃₃₀]	motif	4106	4123	18
Motif: SATPP ₁₃₆₂	motif	4205	4219	15
Motif: VATNIIENGVT ₁₄₆₉	motif	4505	4540	36
Motif: QRLGRVGR ₁₅₁₃	motif	4649	4672	24
CI/6K2: VHHQ ₁₇₉₀ /S	Cleavage site	5492	5506	15
6K2	CDS	5504	5659	156
6K2/VPg: VQHQ ₁₈₄₂ /G	Cleavage site	5648	5662	15
VPg	CDS	5660	6250	591
VPg/N1a: ATHE ₂₀₃₉ /G	Cleavage site	6239	6253	15
N1a-Pro	CDS	6251	6979	729
N1a-Pro/N1b: VKTQ ₂₂₈₂ /G	Cleavage site	6968	6982	15
N1b RdRp	CDS	6980	8536	1557
Motif: SLKAEL ₂₄₅₈	motif	7490	7507	18
Motif III: WTVGMTKIFYGGWNELMRKL ₂₅₂₁	motif	7640	7696	57
Motif IV: VDADGSQFDSSL ₂₅₃₉	motif	7715	7750	36
Motif V: KGNNSGQPSTVVVDNTLMVMLAMYY ₂₆₁₂	motif	7898	7969	72
Motif VI: FYANGDDLII ₂₆₃₉	motif	8021	8050	30
Motif VII: EDLWFMSHQAKV ₂₆₈₄	motif	8150	8185	36
Motif VIII: YIPKLEQERIVS ₂₇₀₀	motif	8198	8233	36
N1b/CP: VSHQ ₂₈₀₁ /A	Cleavage site	8525	8539	15
CP	CDS	8537	9445	909
Motif: DAG ₂₈₁₆	motif	8573	8581	9
3'-UTR	3'UTR	9446	9608	163

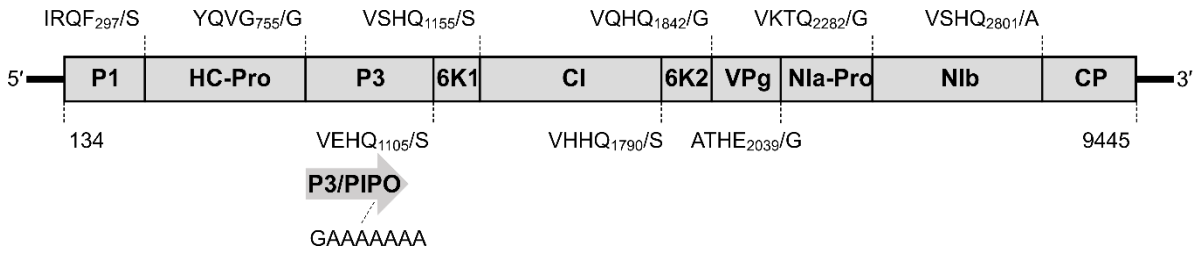


Fig. 3 Genomic organization of YMV and predicted cleavage sites.

3.3. Pairwise sequence analyses of the Brazilian genomes

The sequences of YMV from Brazil (YMV_DR1, YMV_DR2, YMV_Dsp, and YMV_I4) showed ~90-91% nt identity with the complete YMV sequence from Nigeria (GenBank acc. MG711313), and 86% nt identity with the sequence from Ivory Coast (GenBank acc. NC_004752). A detailed pairwise sequence comparison among complete genomic sequences is shown in Table 3. The lowest identity in the polyprotein coding region was observed for the P1, which showed nucleotide identities of ~89% and ~78% with the sequences from Nigeria and Ivory Coast, respectively (Table 3).

Pairwise comparison of the four sequences from Brazil using the “dataset_11” sequences revealed that the partial sequence regions of the P1 shared 97.0 - 99.4% identity at the nt level and 97.5 - 99.6 % at aa level; the HC-Pro region shared 98.0 - 99.8 % nt and 98.7 - 100 % aa identities; the P3 region shared 98.7- 99.7 % nt and 98.7 - 100 % aa identities and the NIb region shared identities of 98.2 - 99.7 % at nt and 99.7 - 100 % at aa levels (Supplementary Fig. S1).

Table 3 Pairwise comparisons of nucleotide and deduced aminoacid sequences of nucleotide genomic sequences, polyproteins, individual coding and non-coding regions of YMV from Brazil (YMV_DR1, YMV_DR2, YMV_Dsp, and YMV_I4) and reference sequences using SDT software.

Isolates		P1		HC-Pro		P3		6K1		CI		6K2		NIa		VPg		Nlb		CP		POLY		5'UTR	3'UTR	WG
		nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	nt	nt
DR1	DR2	0.99	0.99	0.99	1.00	0.99	1.00	0.98	1.00	0.99	1.00	0.99	1.00	0.99	1.00	0.99	1.00	0.99	0.99	0.99	1.00	0.99	1.00	0.98	0.99	0.99
	Dsp	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99	0.98	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99
	I4	0.99	0.98	0.99	0.99	0.98	0.99	0.99	1.00	0.99	1.00	0.99	1.00	0.98	1.00	0.98	1.00	0.99	1.00	0.98	1.00	0.98	1.00	0.98	0.99	0.98
	NIG	0.89	0.88	0.91	0.98	0.90	0.97	0.89	0.98	0.90	0.99	0.95	0.98	0.91	0.99	0.91	0.98	0.91	0.97	0.93	0.95	0.91	0.97	0.87	0.94	0.91
	CDI	0.78	0.76	0.89	0.97	0.85	0.90	0.86	0.98	0.86	0.97	0.81	0.92	0.88	0.97	0.90	0.97	0.88	0.96	0.83	0.85	0.86	0.93	0.83	0.92	0.86
DR2	Dsp	0.98	0.99	0.99	1.00	0.99	0.99	0.98	1.00	0.99	1.00	0.99	1.00	0.99	1.00	0.99	0.98	0.99	0.99	0.99	1.00	0.99	1.00	0.95	0.99	0.99
	I4	0.97	0.98	0.98	0.99	0.98	1.00	0.98	1.00	0.98	1.00	0.98	1.00	0.98	1.00	0.98	1.00	0.98	0.99	0.98	1.00	0.98	0.99	0.98	0.98	0.98
	NIG	0.89	0.88	0.91	0.98	0.90	0.97	0.89	0.98	0.90	0.99	0.94	0.98	0.91	0.99	0.91	0.98	0.90	0.96	0.93	0.95	0.91	0.97	0.87	0.93	0.91
	CDI	0.78	0.76	0.88	0.97	0.85	0.91	0.87	0.98	0.86	0.97	0.80	0.92	0.87	0.97	0.90	0.97	0.88	0.96	0.84	0.85	0.86	0.93	0.85	0.92	0.86
Dsp	I4	0.98	0.98	0.99	0.99	0.98	0.99	0.99	0.98	0.98	1.00	0.99	1.00	0.99	1.00	0.98	0.98	0.99	1.00	0.98	1.00	0.98	0.99	0.95	0.99	0.98
	NIG	0.89	0.88	0.91	0.99	0.90	0.97	0.89	0.98	0.90	0.99	0.95	0.98	0.91	0.99	0.91	0.96	0.91	0.97	0.93	0.95	0.91	0.97	0.85	0.94	0.91
	DCI	0.78	0.76	0.89	0.97	0.85	0.91	0.86	0.98	0.86	0.97	0.81	0.92	0.87	0.97	0.89	0.95	0.88	0.96	0.83	0.85	0.86	0.93	0.83	0.92	0.86
I4	NIG	0.88	0.87	0.90	0.98	0.90	0.97	0.88	0.98	0.90	0.99	0.94	0.98	0.91	0.99	0.90	0.98	0.91	0.97	0.92	0.94	0.90	0.97	0.89	0.93	0.90
	DCI	0.77	0.75	0.88	0.96	0.85	0.91	0.86	0.97	0.86	0.97	0.82	0.92	0.87	0.97	0.89	0.97	0.88	0.96	0.83	0.85	0.86	0.93	0.83	0.91	0.86

Reference sequences of YMV from Ivory Coast (CDI, NCBI accession number: NC_004752), and Nigeria (NIG, NCBI accession number: MG711313) were used for alignments. Polyprotein (POLY), protein 1 (P1), helper component protease (HC-Pro), protein 3 (P3), 6 kilodalton peptide 1 (6K1), 6 kilodalton peptide 2 (6K2), cytoplasmic inclusion (CI), nuclear inclusion A protease (NIa), viral protein genome-linked (VPg), nuclear inclusion B RNA-dependent RNA polymerase (Nlb), coat protein (CP), 5'-untranslated region (5'-UTR), 3'-untranslated region (3'-UTR), whole-genome (WG).

3.4. Bayesian analyses

A Bayesian phylogenetic analyses on the four subsets of dataset_11 showed that the YMV from Brazil grouped into a new independent clade (Fig. 4), closely related to the NIG sequences (group VII) and more distantly related to the BUF sequences (group I) (Fig. 4 and Supplementary Fig. S1). To obtain a better overview of the phylogenetic relationships, the same analysis was performed on “dataset_32_CP”, by comparing YMV from Brazil with isolates from the established nine phylogenetic groups. The results further supported the clustering of YMV from Brazil into a new monophyletic group (group X), distinct from the nine established groups (Fig. 5). Same results were obtained when using the N-terminus region of CP, the CP core/C-terminus and a region including the partial Nib, the complete CP and 3'-UTR regions (data not shown).

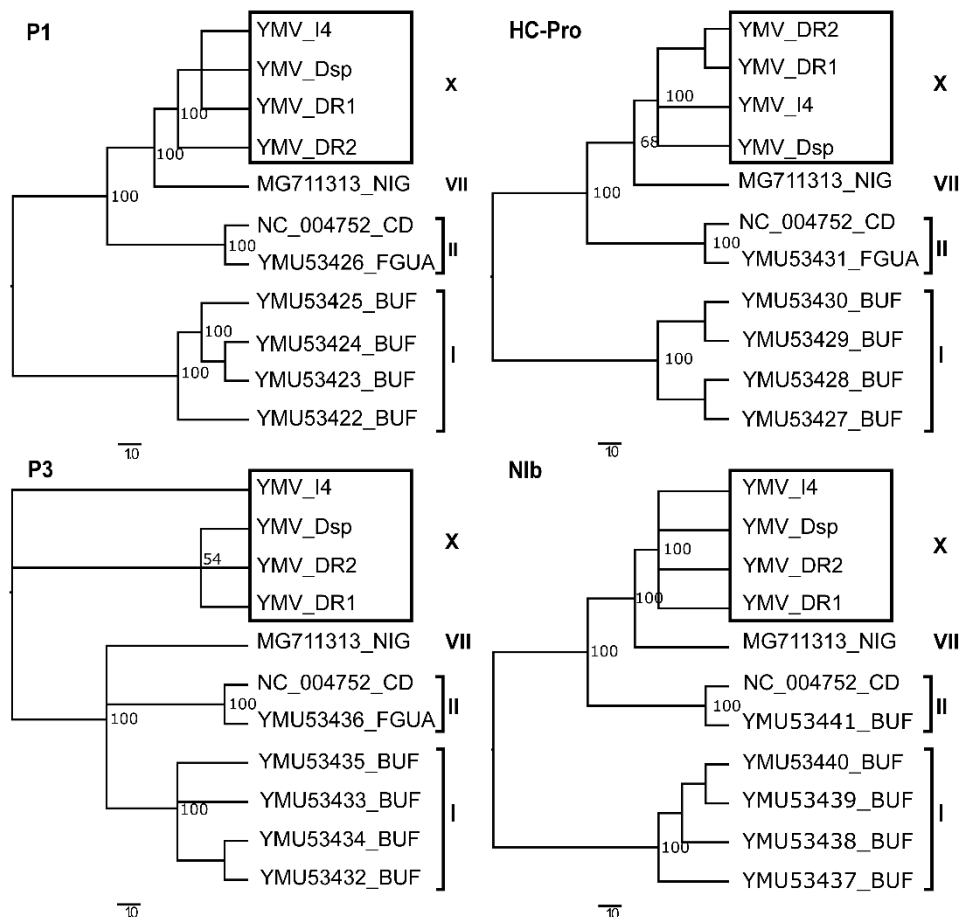
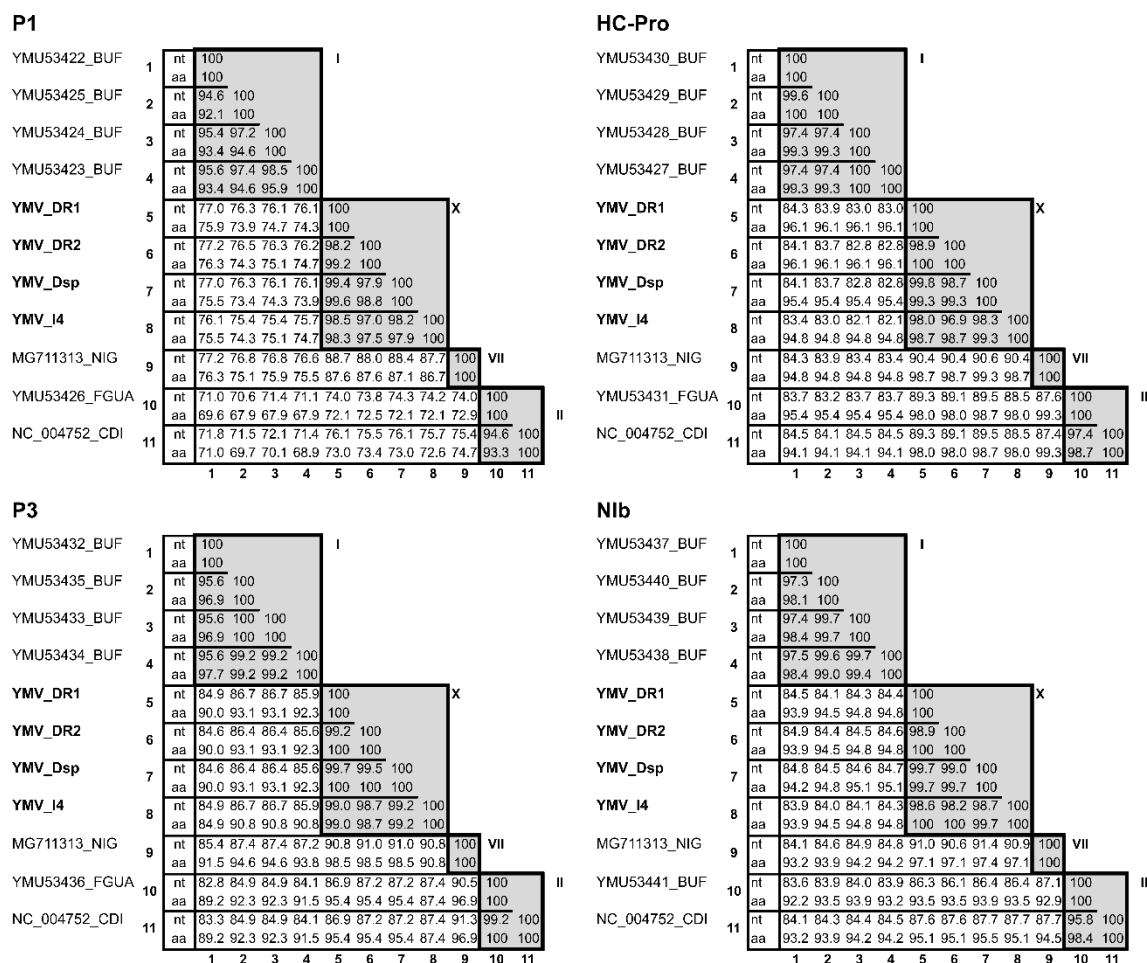


Fig. 4 Phylogenetic trees based on partial nucleotide sequences of P1, HC-Pro, P3, and partial C-Ter regions of Nib of eleven YMV genome sequences using Bayesian inference.



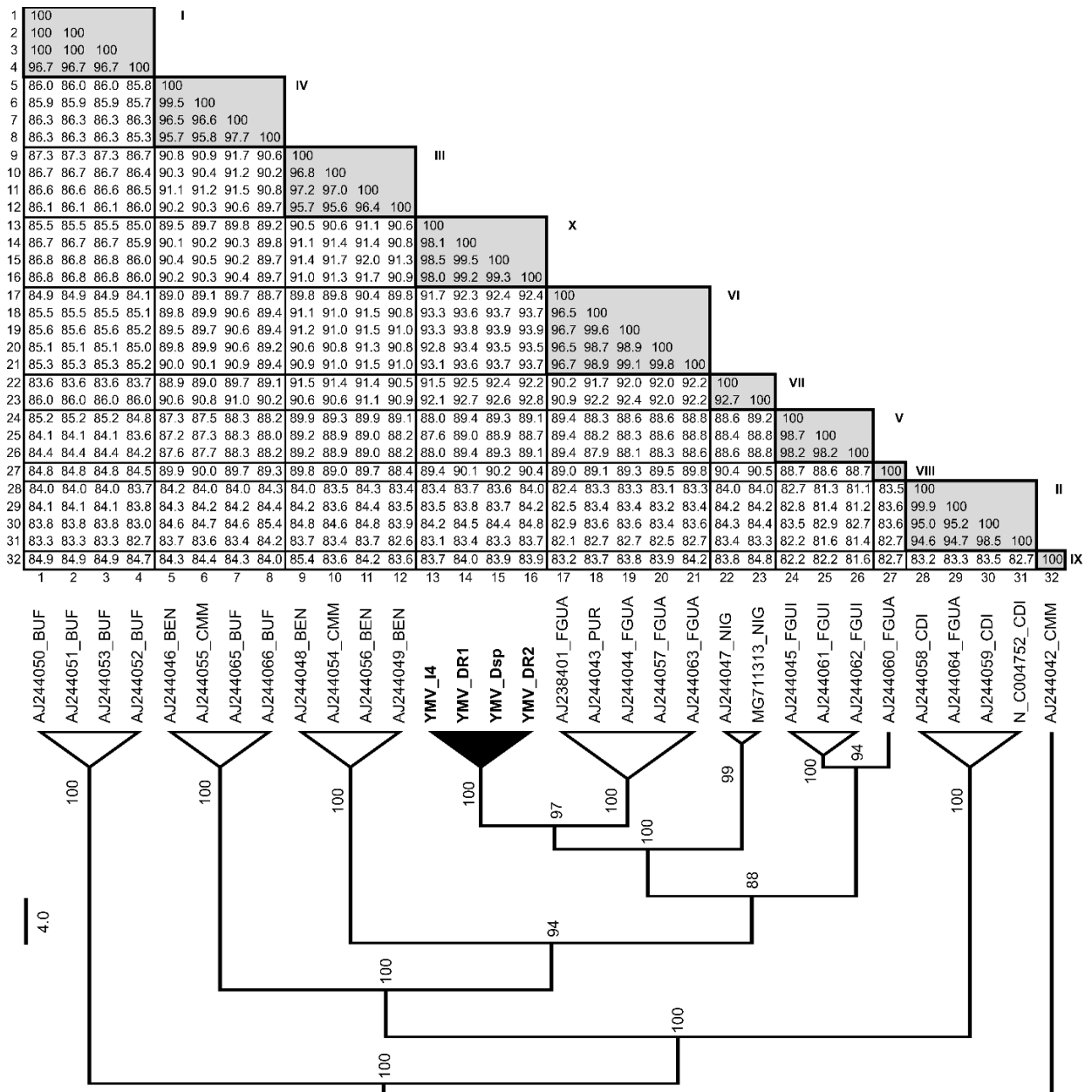


Fig. 5 Bayesian phylogenetic analysis of complete CP YMV sequences. Numbers on the branches represent mean posterior probability values, and the scale bar represents the nucleotide substitutions per site. The matrix reports the pairwise identities inferred with SDT v.1.2 software.

The pairwise comparisons of CP sequences at nt and aa level showed that the Brazilian group shared the highest nt identities with sequences from group VI (91.7 - 93.9 % nt, 92.1 - 95 % aa), which is composed of isolates from Guadeloupe (FGUA) and Puerto Rico (PUR), followed by group VII (91.5 - 92.8 % nt, 94.4 - 95.4 % aa), that is composed of isolates from Nigeria (NIG) and group III (90.4 - 92.0 % nt, 90.7 - 92.4 % aa), having isolates from Benin (BEN) and Cameroon (CMM). In contrast, the lowest identities (83.1 - 84.8 % nt, 85.1 - 86.8 %

334 aa) were observed with group II, composed of isolates from the Ivory Coast (CDI) and
335 Guadeloupe (FGUA) and group IX with an isolate from Cameroon (CMM) (Fig. 5, Table 4).
336

Table 4 Estimations of inter-and intra-groups genetic diversity using CP sequences by nucleotide and amino acid (in parenthesis) identities comparisons between and within YMV isolate groups.

Inter-group	nt/aa	I	II	III	IV	V	VI	VII	VIII ^a	IX ^a	
II	nt	82.7-84.1									
	aa	(83.8-86.6)									
III	nt	86.0-87.3	82.6-84.8								
	aa	(85.9-88.6)	(83.4-87.4)								
IV	nt	85.3-86.3	83.4-85.4	89.7-91.7							
	aa	(86.6-87.9)	(85.1-87.7)	(90.7-93.7)							
V	nt	83.6-85.2	81.1-83.5	88.2-89.9	87.2-88.3						
	aa	(83.3-85.5)	(81.5-84.8)	(86.8-90.4)	(88.1-90.1)						
VI	nt	84.1-85.6	82.1-83.6	89.8-91.5	88.7-90.9	87.9-89.4					
	aa	(84.1-87.2)	(82.5-85.4)	(89.1-93.4)	(89.1-92.7)	(87.4-89.7)					
VII	nt	83.6-86.0	83.3-84.4	90.5-91.5	88.9-91.0	88.4-89.2	90.2-92.4				
	aa	(85.9-86.9)	(84.4-86.1)	(92.4-93.7)	(92.1-93.4)	(88.1-92.1)	(90.7-94.0)				
VIII	nt	84.5-84.8	82.7-83.6	88.4-89.8	89.3-90.0	88.6-88.7	89.1-89.8	90.4-90.5			
	aa	(84.8-85.9)	(82.8-84.8)	(88.7-90.7)	(90.7-91.4)	(85.4-87.1)	(89.4-92.1)	(90.7-91.1)			
IX	nt	84.7-84.9	82.7-83.5	83.6-85.4	84.0-84.4	81.6-82.2	83.2-84.2	83.8-84.8	82.7		
	aa	(84.1)	(84.1-85.1)	(84.1-84.8)	(84.1-85.4)	(81.1-81.8)	(81.1-84.8)	(83.8-84.4)	(82.5)		
X	nt	85.0-86.8	83.1-84.8	90.4-92.0	89.2-90.5	87.6-89.4	91.7-93.9	91.5-92.8	89.4-90.4	83.7-84.0	
	aa	(86.2-87.2)	(85.1-86.8)	(90.7-92.4)	(92.1-93.0)	(89.1-91.1)	(92.1-95)	(94.4-95.4)	(90.7-91.1)	(85.1-85.4)	
		I	II	III	IV	V	VI	VII	VIII ^a	IX ^a	X
Intra-Group	nt	96.7-100	94.6-99.9	95.6-97.2	95.7-99.5	98.2-98.7	96.5-99.8	92.7			98.0-99.5
	aa	(95.9-100)	(93.0-99.7)	(96.0-97.7)	(96.7-99.0)	(97.0-98.3)	(96.0-99.7)	(94.4)			(99.7-100)

^a only one isolate. Group I (isolates from Burkina Faso-BUF), group II (Ivory Coast and Guadeloupe, CDI-FGUA), group III (Benin and Cameroon, BEN-CMM), group IV (Benin, Burkina Faso, and Cameroon, BEN-BUF-CMM), group V (French Guiana, FGUI), group VI Guadeloupe and Puerto Rico, FGUA-PUR), group VII (Nigeria, NIG), group VIII (Guadeloupe, FGUA), group IX (Cameroon, CMM), and group X (sequences from Brazil, BRA).

3.5. Geospatial analysis

Geographic, ecological and phylogenetic data from CP YMV sequences were merged in a single analysis. Phylogenetic projections showed that YMV isolates I (BUF), II (FGUA-CDI), III (BEN-CMM), IV (BEN-BUF-CMM), VII (NIG), IX (CMM), and X (sequences from Brazil) fall in groups infecting *D. cayennensis-rotundata*. YMV isolates from group IV (BEN-BUF-CMM), groups VI (PUR) and VIII (FGUA) also infect *D. alata*. Isolates from these groups come from regions of tropical rain forest and shrubland ecological zones. Group V (FGUI) was composed of three YMV isolates that infect *D. trifida* found in coastal plain regions. Interestingly, the geospatial analysis also revealed that the diversity of YMV isolates is based on geographic origin. Indeed, YMV isolates from groups IV and VI originated from different species of *Dioscorea* (*D. cayennensis-rotundata*, *D. alata*, and *D. trifida*) and clustered with location of those isolates (Fig. 6). Thus, YMV isolates of group IV (BEN-BUF-CMM) of *D. cayennensis-rotundata* and *D. alata*, and of group VI (FGUA-PUR) of *D. cayennensis-rotundata*, *D. alata* and *D. trifida* are clustering according to ecological regions/environment classified as tropical rain forests and not to host.

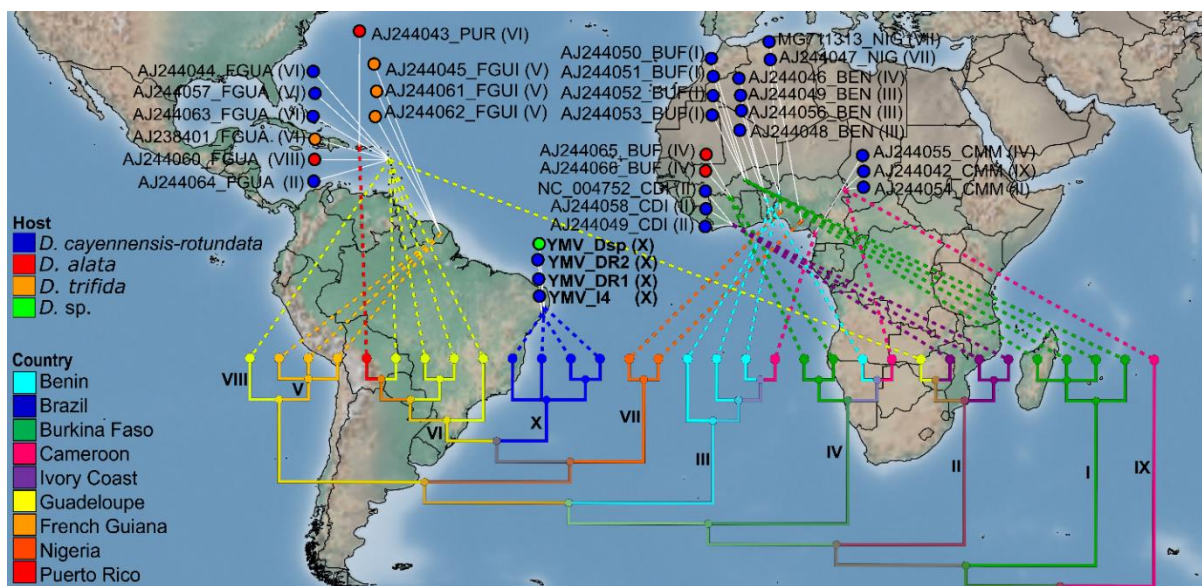


Fig. 6 Geospatial phylogenetic analysis using complete CP sequences of Brazilian YMV (YMV_DR1, YMV_DR2, and YMV_Dsp) and 28 sequences retrieved from NCBI. Circles with bold profile represent the host: blue (*D. cayennensis-rotundata*), red (*D. alata*), orange (*D. trifida*) and green (*D. sp.*). Colors of branches represent the geographic origins of YMV isolates. Phylogroups are represented by the group I (BUF), group II (CDI-FGUA), group III (BEN, CMM), group IV (BEN-BUF-CMM), group V (FGUI), group VI (FGUA-PUR), group VII

(NIG), group VIII (FGUA), group IX (CMM) and the newly determined group X (YMV sequences from Brazil).

3.6. Evolutionary and temporal analysis

Evolutionary and temporal analyses were performed with BEAST. After obtaining a significant phylogenetic signal of the “dataset_26” (Supplementary Table S4) using DAMBE and no evidence of recombination events with RDP, TempEst confirmed that the dataset had a sufficient temporal signal for performing the phylogenetic molecular clock analysis. Root-to-tip linear regression analyses between genetic divergence and sampling date showed that “dataset_26” had a positive temporal signal (correlation coefficient = 0.2831). First, we tested the strict model, but no correlation between the lineage’s rate and its ancestor was found, and therefore, the uncorrelated models were used. The ucl.d.stdev parameter estimated values close to 0.0, indicating the appropriated relaxed molecular clock. As mentioned before, two clock models were assessed under the GTR+G+I substitution model (BIC). The best-fitted likelihood estimations were obtained for the uncorrelated lognormal relaxed clock (UCLN) with exponential distribution (Supplementary Table S6). The mean rate of nucleotide substitutions of YMV CP was estimated to be 1.76×10^{-3} substitutions/site/year (95 % HPD interval: 2.45×10^{-4} - 3.90×10^{-3}). The mean time to the most recent common ancestor (TMRCA) of the tree root was estimated at 168.68 years ago (95 % HPD: 48.56-363.28 years) (Fig. 7), indicating that YMV at least since ~1850 spread from *D. cayennensis-rotundata* to other *Dioscorea* species (Fig. 7 and Supplementary Fig. S2). The spatial-temporal dynamics of dissemination of YMV indicates that YMV migrated from Africa (Nigeria, NIG) to Guadeloupe (FGUA) (Supplementary Fig. S2) around 63 years ago before reaching yam fields in Brazil (Fig. 7). Thus, according to our evolutionary and temporal analysis YMV has an African origin from the yam complex *D. cayennensis-rotundata* followed by independent transfers to *D. trifida* from French Guiana around 80 years ago and more recently to *D. alata* from Burkina Faso and Puerto Rico about 53 and 30 years ago, respectively (Supplementary Fig. S2 and Fig.7).

392 **Supplementary**

393 **Table S6.** Bayesian phylogenetic estimation of evolutionary parameters and model comparison for CP sequences of YMV under different clock
394 models and coalescent tree.

Clock Model: Relaxed Distribution	Tree prior	Substitution rate (substitutions/site/year)			TMRCA			PS*	Rank	SS*	Rank
		Mean	Lower 95 % HPD	Upper 95 % HPD	Mean	Lower 95 % HPD	Upper 95 % HPD				
Lognormal	Exponential Growth	1.76E-03	2.45E-04	3.90E-03	168.68	48.56	363.28	-4735.3881	1	-4724.4946	1
	Expansion Growth	1.85E-03	3.18E-04	3.96E-03	161.77	49.18	339.95	-4736.0273	2	-4725.5115	2
	Bayesian Skyline	9.76E-04	1.17E-04	2.37E-03	283.54	49.00	636.39	-4748.2769	3	-4748.4512	4
	Constant Size	8.10E-04	3.11E-05	1.94E-03	415.80	59.86	1045.71	-4785.4017	7	-4787.7465	7
Exponential	Exponential Growth	4.88E-03	2.91E-03	7.09E-03	63.78	47.01	84.40	-4754.7773	5	-4754.8269	6
	Expansion Growth	7.38E-03	4.72E-03	1.04E-02	52.07	42.00	63.48	-4748.6649	4	-4742.7481	3
	Bayesian Skyline	1.29E-02	9.58E-03	1.60E-02	42.39	42.00	43.60	-4800.3526	8	-4800.8341	8
	Constant Size	7.03E-03	4.52E-03	9.93E-03	53.58	42.33	66.91	-4756.4069	6	-4753.7052	5

395 **PS:** Path Sampling, **SS:** Stepping-stone Sampling.
396 *Mean marginal likelihood of four independent runs.

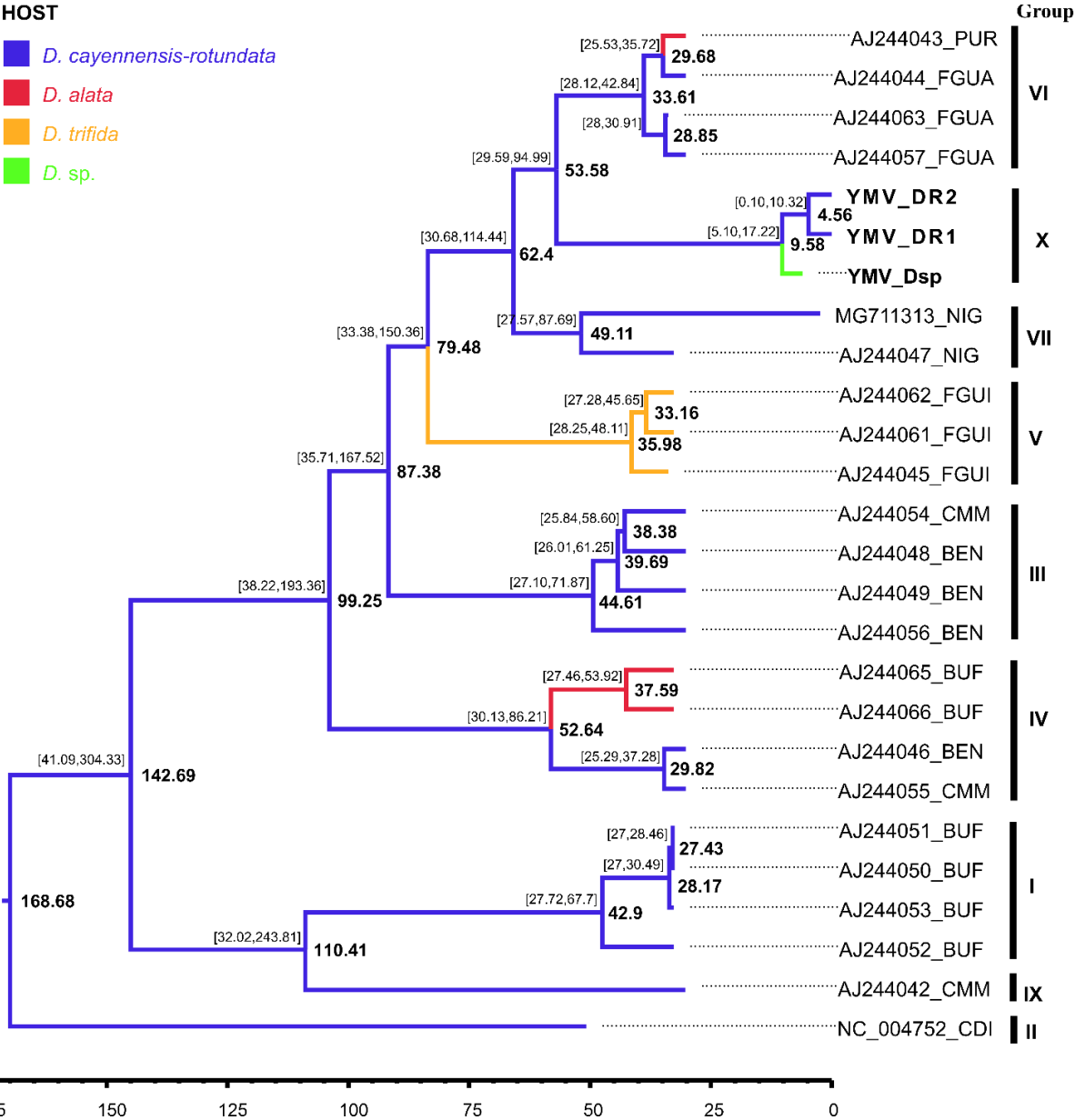
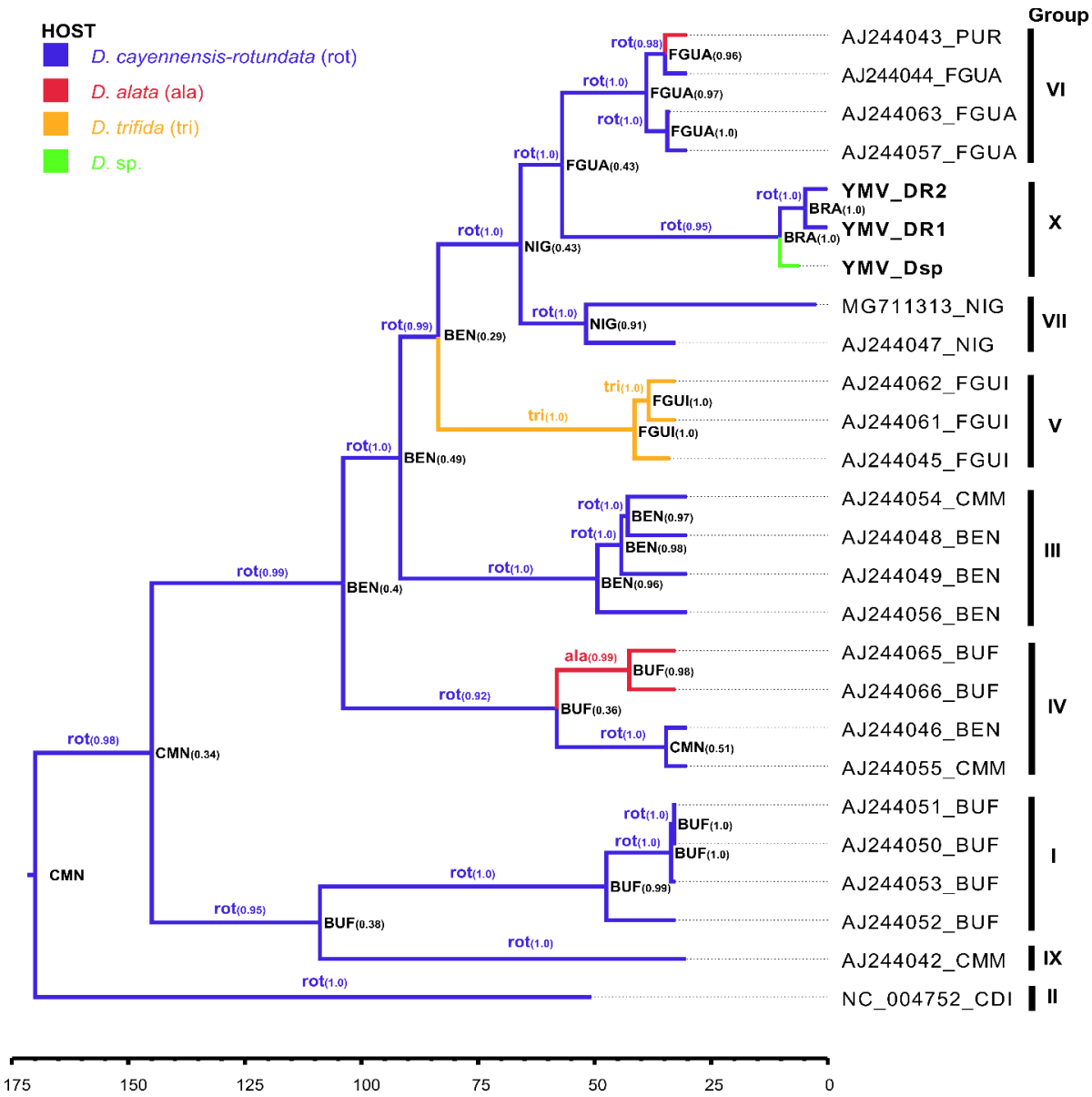


Fig. 7 Maximum Clade Credibility (MCC) tree of CP YMV sequences (“dataset_26”) from 1977 to 2019. Nodes are labeled with the mean height and numbers above branches denote the interval of highest posterior densities (HPD). Scale bars represent years of evolutionary rate.



Supplementary

Fig. S2. Maximum Clade Credibility (MCC) tree of CP YMV sequences (“dataset_26”) from 1977 to 2019. Nodes are labeled with Country labels: Benin (BEN), Brazil (BRA), Burkina Faso (BUF), Ivory Coast (CDI), Cameroon (CMM), Guadeloupe (FGUA), French Guiana (FGUI), Nigeria (NIG), and Puerto Rico (PUR). Right side of the country and host labels in parentheses their respective mean posterior probability values. Scale bars represent years of evolutionary rate.

4. Discussion

In this study, virus surveys were conducted to assess incidence and distribution of YMV in yam growing areas of Paraíba, Pernambuco and Alagoas in Brazil. YMV was common in all regions but was found in *D. cayennensis-rotundata* only. YMMV was also found but restricted

to *D. alata* and there never were mixed infections of these two viruses in *D. cayennensis-rotundata* and *D. alata*. Similar results were observed by Mumford and Seal [51], where YMV was detected via IC-RT-PCR only in *D. cayennensis-rotundata* plants and YMMV only in *D. alata* plants from Africa and India. Furthermore, Umber et al. [52] by indexing an *in vitro* collection of yams at the Biological Resources Center for Tropical Plants (BRC-TP) of Guadeloupe found that accessions of *D. alata* were mostly infected with YMMV, while accessions of *D. cayennensis-rotundata* were mainly infected with YMV. In contrast, Eni et al. [53] detected mixed YMV and YMMV infections in both *D. cayennensis*, *D. rotundata* and *D. alata* plants from Ghana and Togo. Other studies show YMV infecting *D. alata* plants from Guadeloupe, Burkina Faso and Puerto Rico as well *D. trifida* in French Guiana and Guadeloupe [18, 20].

The first full-genomic sequence of YMV was published in 1996 from an isolate collected in 1977 from naturally-infected yam in Ivory Coast and consisted of 9,609 nt (NC_004752) [17]. It was only in 2019 that the second complete genome sequence of a YMV isolate (9,594 nt, YMV-Nigeria, MG711313) became available [7]. In this work, four genomic sequences of YMV infecting yams in Brazil were assembled from HTS and Sanger data, each one 9,608 nt in size. The Bayesian phylogenetic assigned the YMV from Brazil to a new, independent clade (defined as group X) and revealed a high intergroup nucleotide diversity. Several studies have also already reported the relatively high intraspecific genetic variability of YMV isolates of *D. cayennensis-rotundata* (nt identities of 92.7-100% for CP sequences), which was attributed to the differential accumulation of mutations and a significant contribution of recombination events [18, 20]. Surprisingly, we found a relatively low genetic diversity among the Brazilian YMV sequences (with nt identity of 98.4-99.5% for genomic sequences and of 98.2-99.2% for CP sequences). The nucleotide diversity of viral populations has been reported to be influenced by the mode of transmission [54]. As example, in the case of potato virus Y (PVY) it has been found that aphid-borne lineages were the least diverse, while tuber-borne lineages were the more diverse [54]. Transmission bottlenecks can indeed affect viral diversity by drastically limiting the genetic diversity between hosts and appear to be primarily a stochastic mechanism [55]. Moury et al. [56] showed that an aphid transmits an average of 0.5 to 3.2 PVY particles, which is extremely low compared to a viral population within a plant [57], imposing a transmission bottleneck. However, potyvirus transmission by aphids plays a role in dispersal mainly over short distances [58], and therefore, the relatively low intraspecific genetic diversity found for YMV in Brazil remains to be elucidated.

In former phylogenetic analyses, Bousalem and colleagues [20] analyzed the C-terminus of the replicase (N1b), the CP and 3'-UTR regions of 27 YMV isolates to classify them into nine phylogroups: group I (BUF), group II (CDI-FGUA), group III (BEN, CMM), group IV (BEN-BUF-CMM), group V (FGUI), group VI (FGUA-PUR), group VII (NIG), group VIII (FGUA) and group IX (CMM). In our study, the updated analysis of 32 sequences including the sequences from Brazil showed the formation of a new phylogroup (Brazilian group X) that was most closely related to groups VI (FGUA and PUR) and VII (NIG), followed by group V (FGUI). In agreement with these results, the phylogeographic analysis showed that the most probable ancestors of the YMV isolates from Brazil were located in Guadeloupe and Nigeria, with *D. cayennensis-rotundata* as host. Further, the geospatial and evolutionary results showed a relation between the genetic diversity of YMV and geographic location rather than to host.

The phylogeography of the YMV dataset and the demographic history were studied using BEAST [59], that allowed us to explicitly handle the time-structured sequence data. First, the linear regression of root-to-tip distances against collected date implemented in TempEst [32] was used to investigate the temporal structure and the correlation between genetic distance against sampling time. Inappropriate sequences were discarded, and thus a total of 26 sequences including all those available in NCBI were employed for the Bayesian analysis. As more sequences are available, there may be a greater variation in the rate between lineages to justify the use of a relaxed clock model. The implemented methods in BEAST differ in how to deal with the phylogenetic uncertainty and the assumptions about rate variation between lineages. To estimate the evolutionary rate, relaxed clock models that allows different rates along each branch in the tree [39] were employed. Branching rates can be handled as independent and identically distributed samples from a logarithmic normal distribution in the uncorrelated lognormal (UCLN) model or an exponential distribution in uncorrelated exponential (UCEXP) relaxed-clock model [39]. Testing different clock models gave us information on the factors that affect the reliability for estimating rates from the time-structured sequences [60]. In this study, we compared a relaxed clock with a log-normal relaxed distribution and a relaxed clock with exponential distribution models for calculating evolutionary rates of YMV CP genes. In addition, four demographic models (expansion, exponential, Bayesian skyline, and Constant Size) were included in the analyses. Also of notice, we did test if randomizing collected dates and traits would give us the same temporal information contained in the dataset. This randomization allowed us to verify if probabilities were biased. Bayesian approaches have been previously used to estimate rates of evolution of plant viruses. Gibbs et al. [61] used the partial CP of about 50 potyviruses for evolutionary studies and showed that the relaxed clock model

was favored over a strict model by clock likelihood ratio tests. They determined that CP genes of potyviruses have an evolutionary rate of about 1.15×10^{-4} nucleotide substitutions/site/year (nt/s/yr) and the initial radiation of potyviruses occurred about 6,600 years ago. In addition, Mao et al. [62] showed that an uncorrelated lognormal relaxed clock with a Bayesian skyline coalescent tree prior was favored for studying the PVY VPg gene sequences. An evolutionary rate of 5.60×10^{-4} nt/s/yr (95 % HPD interval: $3.35 \times 10^{-4} - 8.17 \times 10^{-4}$) was determined, equivalent to those of other plant RNA viruses. Simmon et al. [63] suggested the coalescence model with an exponential growth of the population under a relaxed molecular clock for the best supported evolutionary model of the ZYMV CP gene and calculated a mean evolutionary rate of 5.0×10^{-4} nt/s/yr (95 % HPD interval: $1.8 - 8.8 \times 10^{-4}$). In this study, BEAST favored the relaxed uncorrelated lognormal distribution with an exponential growth tree prior model with growth rate parameterization for estimating rates of 1.76×10^{-3} nt/s/yr (95 % HPD interval: $2.45 \times 10^{-4} - 3.90 \times 10^{-3}$). We also estimated dN/dS ratios using the renaissance counting method [64] implemented in BEAST. Two independent MCMC runs of 200 million steps were calculated for two runs (data not shown). The dN/dS relation of each specific site was estimated through the posterior distribution of this relation for all the sites with their 95% credibility intervals (CI), considering as positive selection the posterior probability values of $dN/dS > 1$. The evolutionary rate estimated using the CP YMV gene was 9.0816×10^{-4} nt/s/year (95% HPD interval: $1.5381 \times 10^{-4}, 1.8926 \times 10^{-3}$). From the 303 codons only 12 sites showed values dN/dS ratios greater than one showing a negative selection for most of the sites. The obtained dN/dS mean ratio of 0.26 for the CP is similar to what has been reported for other potyviruses [65].

Our analysis dated the initial radiation of the YMV CP gene from the African continent around 168.68 years before present (YBP) (95 % HDP: 48.56 – 363.28). Similar analyses to reveal the probable route of transport were previously performed for papaya ringspot virus (PRSV), which is thought to be transported from India to South America as early as 450 YBP and more likely only 300 years ago [63]. It is very likely that YMV was transported from the African continent to Central America and from there to South America with trading of Europeans reaching Havana as a major port close to Puerto Rico and Guadeloupe [66]. According to our phylogeographic analysis the most recent common ancestor of YMV emerged in Cameroon approximately in 1850 from the yam complex *D. cayennensis-rotundata* and was independently transferred to *D. trifida* from French Guiana around 80 years ago and more recently to *D. alata* from Burkina Faso and Puerto Rico about 53 and 30 years ago, respectively.

Overall, this study reveals the presence of a new phylogroup of YMV from Brazil and contributes key aspects to the probable epidemiological history of YMV.

Acknowledgments

The authors acknowledge the financial support given by CNPq (Project 408709/2018-9), CAPES-PrInt-UFRPE, and the Rural Federal University of Pernambuco (Edital 015/2018, UFRPE-PRPPG).

Thanks to the yam farmers who allowed access to their fields to collect samples of study material.

Also, to professionals, for providing samples (yam tubers) and support in the sample collections in non-cultivated areas: Master in Plant Biology, Gabriel Pavan Sabino, Universidade Estadual Paulista (UNESP), Rio Claro, São Paulo. Botanical Biologist Dr. Diogo Amorim de Araújo, Pernambuco. Botanical Biologist Dr. Jones Clebson R. Mendez, Universidade Federal Rural de Pernambuco (UFRPE), Plant Taxonomy Laboratory, Pernambuco.

Thanks to the researchers of the Agronomic Institute of Pernambuco (IPA) Dr. José Nildo Tabosa, and MSc. Almir Dias Alves da Silva for providing us with access to Yam's production fields.

Funding

This work was supported by funding from CNPq (Project 408709/2018-9), CAPES-PrInt-UFRPE, and the Rural Federal University of Pernambuco (Edital 015/2018, UFRPE-PRPPG).

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Consent for publication All contributing authors have read and approved of the final version of the manuscript.

References

1. Zappi DC, Ranzato Filardi FL, Leitman P, et al (2015) Growing knowledge: An overview of Seed Plant diversity in Brazil. *Rodriguesia* 66:1085–1113. <https://doi.org/10.1590/2175-7860201566411>
2. Siqueira M (2011) Yam: a neglected and underutilized crop in Brazil. *Hortic Bras* 29:16–20. <https://doi.org/10.1590/s0102-05362011000100003>
3. Siqueira M, Nascimento W, Silva L, et al (2014) Distribution, management and diversity

- of yam local varieties in Brazil: a study on *Dioscorea alata* L. *Brazilian J Biol* 74:52–61.
<https://doi.org/10.1590/1519-6984.17112>
4. Kirizawa M, Xifreda CC, Silva JH da (2016) Diversidade florística de Dioscoreaceae na Reserva Biológica do Alto da Serra de Paranapiacaba, Santo André, São Paulo, Brasil. *Hoehnea* 43:99–117. <https://doi.org/10.1590/2236-8906-55/2015>
 5. Bousalem M, Durand O, Scarcelli N, et al (2009) Dilemmas caused by endogenous pararetroviruses regarding the taxonomy and diagnosis of yam (*Dioscorea* spp.) badnaviruses: Analyses to support safe germplasm movement. *Arch Virol* 154:297–314. <https://doi.org/10.1007/s00705-009-0311-2>
 6. Bömer M, Rathnayake AI, Visendi P, et al (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. *Arch Virol* 163:533–538. <https://doi.org/10.1007/s00705-017-3605-9>
 7. Bömer M, Rathnayake AI, Visendi P, et al (2019) Tissue culture and next-generation sequencing: A combined approach for detecting yam (*Dioscorea* spp.) viruses. *Physiol Mol Plant Pathol* 105:54–66. <https://doi.org/10.1016/j.pmpp.2018.06.003>
 8. Mambole IA, Bonheur L, Dumas LS, et al (2014) Molecular characterization of yam virus X, a new potexvirus infecting yams (*Dioscorea* spp) and evidence for the existence of at least three distinct potexviruses infecting yams. *Arch Virol* 159:3421–3426. <https://doi.org/10.1007/s00705-014-2211-3>
 9. Seal S, Turaki A, Muller E, et al (2014) The prevalence of badnaviruses in West African yams (*Dioscorea cayenensis-rotundata*) and evidence of endogenous pararetrovirus sequences in their genomes. *Virus Res* 186:144–154. <https://doi.org/10.1016/j.virusres.2014.01.007>
 10. Menzel W, Thottappilly G, Winter S (2014) Characterization of an isometric virus isolated from yam (*Dioscorea rotundata*) in Nigeria suggests that it belongs to a new species in the genus Aureusvirus. *Arch Virol* 159:603–606. <https://doi.org/10.1007/s00705-013-1845-x>
 11. Bakayoko Y, Kouakou AM, Kouassi AB, et al (2021) Detection and diversity of viruses infecting African yam (*Dioscorea rotundata*) in a collection and F1 progenies in Côte d'Ivoire shed light to plant-to-plant viral transmission. *Plant Pathol* 70:1486–1495. <https://doi.org/10.1111/ppa.13393>
 12. Kenyon L, Shoyinka SA, Hughes J d'A., Odu BO (2001) An overview of viruses infecting *Dioscorea* yams in sub-Saharan Africa. *Plant Virol Sub-Saharan Africa* 432–439
 13. Amusa NA, Adegbite AA, Muhammed S, Baiyewu RA (2003) Yam diseases and its management in Nigeria. *African J Biotechnol* 2:514–525. <https://doi.org/10.5897/ajb2003.000-1099>
 14. Silva G, Oyekanmi J, Nkere CK, et al (2018) Rapid detection of potyviruses from crude plant extracts. *Anal Biochem* 546:17–22. <https://doi.org/10.1016/j.ab.2018.01.019>
 15. Asala SW, Alegbejo MD (2016) Effects of serial planting of seed yam tubers on virus incidence and yam tuber degeneration. *African Crop Sci J* 24:341. <https://doi.org/10.4314/acsj.v24i4.1>

- 592 16. Wylie SJ, Adams M, Chalam C, et al (2017) ICTV virus taxonomy profile: Potyviridae.
593 J Gen Virol 98:352–354. <https://doi.org/10.1099/jgv.0.000740>
- 594 17. Aleman ME, Marcos JF, Brugidou C, et al (1996) The complete nucleotide sequence of
595 yam mosaic virus (Ivory Coast isolate) genomic RNA. Arch Virol 141:1259–1278.
596 <https://doi.org/10.1007/BF01718829>
- 597 18. Aleman-Verdaguer ME, Goudou-Urbino C, Dubern J, et al (1997) Analysis of the
598 sequence diversity of the P1, HC, P3, NIb and CP genomic regions of several yam mosaic
599 potyvirus isolates: Implications for the intraspecies molecular diversity of potyviruses. J
600 Gen Virol 78:1253–1264. <https://doi.org/10.1099/0022-1317-78-6-1253>
- 601 19. Duterme O, Colinet D, Kummert J, Lepoivre P (1996) Determination of the taxonomic
602 position and characterization of yam mosaic virus isolates based on sequence data of the
603 5'-terminal part of the coat protein cistron. Arch Virol 141:1067–1075.
604 <https://doi.org/10.1007/BF01718610>
- 605 20. Bousalem M, Douzery EJP, Fargette D (2000) High genetic diversity, distant
606 phylogenetic relationships and intraspecies recombination events among natural
607 populations of Yam mosaic virus: A contribution to understanding potyvirus evolution.
608 J Gen Virol 81:243–255. <https://doi.org/10.1099/0022-1317-81-1-243>
- 609 21. Kitajima EW (2020) An annotated list of plant viruses and viroids described in Brazil
610 (1926-2018). Biota Neotrop 20:. <https://doi.org/10.1590/1676-0611-bn-2019-0932>
- 611 22. Hayashi EAI, Blawid R, de Melo FL, et al (2017) Complete genome sequence of a
612 putative new secovirus infecting yam (Dioscorea) plants. Arch Virol 162:317–319.
613 <https://doi.org/10.1007/s00705-016-3104-4>
- 614 23. Blawid R, Silva JMF, Nagata T (2017) Discovering and sequencing new plant viral
615 genomes by next-generation sequencing: description of a practical pipeline. Ann Appl
616 Biol 170:301–314. <https://doi.org/10.1111/aab.12345>
- 617 24. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina
618 sequence data. Bioinformatics 30:2114–2120.
619 <https://doi.org/10.1093/bioinformatics/btu170>
- 620 25. Bankevich A, Nurk S, Antipov D, et al (2012) SPAdes: A new genome assembly
621 algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477.
622 <https://doi.org/10.1089/cmb.2012.0021>
- 623 26. Muhire BM, Varsani A, Martin DP (2014) SDT: A virus classification tool based on
624 pairwise sequence alignment and identity calculation. PLoS One 9:.
625 <https://doi.org/10.1371/journal.pone.0108277>
- 626 27. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7:
627 Improvements in performance and usability. Mol Biol Evol 30:772–780.
628 <https://doi.org/10.1093/molbev/mst010>
- 629 28. Darriba D, Taboada GL, Doallo R, Posada D (2012) JModelTest 2: More models, new
630 heuristics and parallel computing. Nat Methods 9:772.
631 <https://doi.org/10.1038/nmeth.2109>
- 632 29. Ronquist F, Teslenko M, Van Der Mark P, et al (2012) MrBayes 3.2: Efficient bayesian
633 phylogenetic inference and model choice across a large model space. Syst Biol 61:539–
634 542. <https://doi.org/10.1093/sysbio/sys029>

- 635 30. Rambaut A, Drummond AJ, Xie D, et al (2018) Posterior summarization in Bayesian
636 phylogenetics using Tracer 1.7. *Syst Biol* 67:901–904.
637 <https://doi.org/10.1093/sysbio/syy032>
- 638 31. Parks DH, Mankowski T, Zangooei S, et al (2013) GenGIS 2: Geospatial Analysis of
639 Traditional and Genetic Biodiversity, with New Gradient Algorithms and an Extensible
640 Plugin Framework. *PLoS One* 8:. <https://doi.org/10.1371/journal.pone.0069885>
- 641 32. Rambaut A, Lam TT, Carvalho LM, Pybus OG (2016) Exploring the temporal structure
642 of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2:1–7.
643 <https://doi.org/10.1093/ve/vew007>
- 644 33. Xia X (2017) DAMBE6: New tools for microbial genomics, phylogenetics, and
645 molecular evolution. *J Hered* 108:431–437. <https://doi.org/10.1093/jhered/esx033>
- 646 34. Kumar S, Stecher G, Li M, et al (2018) MEGA X: Molecular evolutionary genetics
647 analysis across computing platforms. *Mol Biol Evol* 35:1547–1549.
648 <https://doi.org/10.1093/molbev/msy096>
- 649 35. Martin DP, Murrell B, Golden M, et al (2015) RDP4: Detection and analysis of
650 recombination patterns in virus genomes. *Virus Evol* 1:1–5.
651 <https://doi.org/10.1093/ve/vev003>
- 652 36. Suchard MA, Lemey P, Baele G, et al (2018) Bayesian phylogenetic and phylodynamic
653 data integration using BEAST 1.10. *Virus Evol* 4:1–5. <https://doi.org/10.1093/ve/vey016>
- 654 37. Suchard MA, Rambaut A (2009) Many-core algorithms for statistical phylogenetics.
655 *Bioinformatics* 25:1370–1376. <https://doi.org/10.1093/bioinformatics/btp244>
- 656 38. Duchêne S, Duchêne D, Holmes EC, Ho SYW (2015) The performance of the date-
657 randomization test in phylogenetic analyses of time-structured virus data. *Mol Biol Evol*
658 32:1895–1906. <https://doi.org/10.1093/molbev/msv056>
- 659 39. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and
660 dating with confidence. *PLoS Biol* 4:699–710.
661 <https://doi.org/10.1371/journal.pbio.0040088>
- 662 40. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference
663 of past population dynamics from molecular sequences. *Mol Biol Evol* 22:1185–1192.
664 <https://doi.org/10.1093/molbev/msi103>
- 665 41. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation
666 parameters, population history and genealogy simultaneously from temporally spaced
667 sequence data. *Genetics* 161:1307–1320. <https://doi.org/10.1093/genetics/161.3.1307>
- 668 42. Griffiths RC, Tavaré S (1994) Sampling theory for neutral alleles in a varying
669 environment. *Philos Trans R Soc Lond B Biol Sci* 344:403–410.
670 <https://doi.org/10.1098/rstb.1994.0079>
- 671 43. Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography
672 finds its roots. *PLoS Comput Biol* 5:. <https://doi.org/10.1371/journal.pcbi.1000520>
- 673 44. Gelman A, Meng XL (1998) Simulating normalizing constants: From importance
674 sampling to bridge sampling to path sampling. *Stat Sci* 13:163–185.
675 <https://doi.org/10.1214/ss/1028905934>
- 676 45. Baele G, Li WLS, Drummond AJ, et al (2013) Accurate model selection of relaxed

- 677 molecular clocks in Bayesian phylogenetics. *Mol Biol Evol* 30:239–243.
678 <https://doi.org/10.1093/molbev/mss243>
- 679 46. Shibolet Y, Haronsky E, Leibman D, et al (2007) The Conserved FRNK Box in HC-
680 Pro, a Plant Viral Suppressor of Gene Silencing, Is Required for Small RNA Binding
681 and Mediates Symptom Development. *J Virol* 81:13135–13148.
682 <https://doi.org/10.1128/jvi.01031-07>
- 683 47. Kadaré G, Haenni A-L (1997) MINIREVIEW Virus-Encoded RNA Helicases. *J Virol*
684 71:2583–2590
- 685 48. Mikami A, Paschal BM, Mazumdar M, Vallee RB (1993) Molecular cloning of the
686 retrograde transport motor cytoplasmic dynein (MAP 1C). *Neuron* 10:787–796.
687 [https://doi.org/10.1016/0896-6273\(93\)90195-W](https://doi.org/10.1016/0896-6273(93)90195-W)
- 688 49. Fernández A, Guo HS, Sáenz P, et al (1997) The motif V of plum pox potyvirus CI RNA
689 helicase is involved in NTP hydrolysis and is essential for virus RNA replication. *Nucleic*
690 *Acids Res* 25:4474–4480. <https://doi.org/10.1093/nar/25.22.4474>
- 691 50. Moradi Z, Mehrvar M, Nazifi E, Zakiaghl M (2017) Iranian johnsongrass mosaic virus:
692 the complete genome sequence, molecular and biological characterization, and
693 comparison of coat protein gene sequences. *Virus Genes* 53:77–88.
694 <https://doi.org/10.1007/s11262-016-1389-8>
- 695 51. Mumford RA, Seal SE (1997) Rapid single-tube immunocapture RT-PCR for the
696 detection of two yam potyviruses. *J Virol Methods* 69:73–79.
697 [https://doi.org/10.1016/S0166-0934\(97\)00141-9](https://doi.org/10.1016/S0166-0934(97)00141-9)
- 698 52. Umber M, Filloux D, Gélabale S, et al (2020) Molecular viral diagnosis and sanitation
699 of yam genetic resources: Implications for safe yam germplasm exchange. *Viruses* 12:.
700 <https://doi.org/10.3390/v12101101>
- 701 53. Eni AO, Hughes JDA, Asiedu R, Rey MEC (2010) Survey of the incidence and
702 distribution of viruses infecting yam (*Dioscorea* spp.) in Ghana and Togo. *Ann Appl Biol*
703 156:243–251. <https://doi.org/10.1111/j.1744-7348.2009.00383.x>
- 704 54. da Silva W, Kutnjak D, Xu Y, et al (2020) Transmission modes affect the population
705 structure of potato virus Y in potato. *PLoS Pathog* 16:1–23.
706 <https://doi.org/10.1371/journal.ppat.1008608>
- 707 55. McCrone JT, Llaure AS (2018) Genetic bottlenecks in intraspecies virus transmission.
708 *Curr Opin Virol* 28:20–25. <https://doi.org/10.1016/j.coviro.2017.10.008>
- 709 56. Moury B, Fabre F, Senoussi R (2007) Estimation of the number of virus particles
710 transmitted by an insect vector. *Proc Natl Acad Sci U S A* 104:17891–17896.
711 <https://doi.org/10.1073/pnas.0702739104>
- 712 57. Malpica JM, Fraile A, Moreno I, Obies CI, Drake JW G-AF (2002) The Rate and
713 Character of Spontaneous Mutation in an RNA Virus. *Genetics* 162:1505–1511.
714 <https://doi.org/10.1093/genetics/162.4.1505>
- 715 58. Pleydell DRJ, Soubeyrand S, Dallot S, et al (2018) Estimation of the dispersal distances
716 of an aphid-borne virus in a patchy landscape. *PLoS Comput Biol* 14:1–24.
717 <https://doi.org/10.1371/journal.pcbi.1006085>
- 718 59. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with
719 BEAUti and the BEAST 1.7. *Mol Biol Evol* 29:1969–1973.

- 720 <https://doi.org/10.1093/molbev/mss075>
- 721 60. Duchêne S, Geoghegan JL, Holmes EC, Ho SYW (2016) Estimating evolutionary rates
722 using time-structured data: A general comparison of phylogenetic methods.
723 *Bioinformatics* 32:3375–3379. <https://doi.org/10.1093/bioinformatics/btw421>
- 724 61. Gibbs AJ, Ohshima K, Phillips MJ, Gibbs MJ (2008) The prehistory of potyviruses:
725 Their initial radiation was during the dawn of agriculture. *PLoS One* 3:
726 <https://doi.org/10.1371/journal.pone.0002523>
- 727 62. Mao Y, Sun X, Shen J, et al (2019) Molecular evolutionary analysis of potato virus Y
728 infecting potato based on the VPg gene. *Front Microbiol* 10:1–11.
729 <https://doi.org/10.3389/fmicb.2019.01708>
- 730 63. Simmons HE, Holmes EC, Stephenson AG (2008) Rapid evolutionary dynamics of
731 zucchini yellow mosaic virus. *J Gen Virol* 89:1081–1085.
732 <https://doi.org/10.1099/vir.0.83543-0>
- 733 64. Lemey P, Minin VN, Bielejec F, et al (2012) A counting renaissance: Combining
734 stochastic mapping and empirical Bayes to quickly detect amino acid sites under positive
735 selection. *Bioinformatics* 28:3248–3256. <https://doi.org/10.1093/bioinformatics/bts580>
- 736 65. Gibbs AJ, Hajizadeh M, Ohshima K, Jones RAC (2020) The potyviruses: An
737 evolutionary synthesis is emerging. *Viruses* 12:1–30. <https://doi.org/10.3390/v12020132>
- 738 66. Moya JC (2018) Migration and the historical formation of Latin America in a global
739 perspective. *Sociologias* 20(49):24–68.
740 <https://doi.org/https://doi.org/10.1590/15174522-02004902>
- 741

CAPÍTULO III

**Near-complete genomic sequences of yam mild mosaic viruses of *D. alata* from
Brazil**

Near-complete genomic sequences of yam mild mosaic viruses of *D. alata* from Brazil

Alejandro Risco Mendoza^{1,2}, Carlos Henrique Machado Dias de Brito¹, Tatsuya Nagata³,
Stephan Winter⁴, Rosana Blawid¹

¹ Department of Agronomy, Fitossanidade, Laboratory of Plant Virology, Federal Rural University of Pernambuco, Brazil, alejandro.mendoza@ufrpe.br, henriquemachadoufrpe@gmail.com, rosana.blawid@ufrpe.br

² Department of Plant Pathology, Agronomy faculty, Universidad Nacional Agraria La Molina, Perú, alejandrorisco@lamolina.edu.pe

³ Department of Cellular Biology, University of Brasília, Campus Darcy Ribeiro, Brasília-DF, Brazil, tatsuya@unb.br

⁴ Plant Virus Department, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH, Germany, stephan.winter@dsmz.de, paolo.margaria@dsmz.de

Abstract

Yam mild mosaic virus (YMMV) is a member of the *Potyvirus* genus that mainly infects yams (*Dioscorea* spp.). In this study, total RNA extractions from two plant groups of *D. alata* (Pool DA) and *Dioscorea* spp. (pool Dsp) was employed for High-Throughput Sequencing (HTS). Two YMMV nucleotide genomic sequences were assembled consisting of 9515 nt (YMMV_DA) and 9501 nt (YMMV_Dsp) in length excluding poly-(A) tails. YMMV_DA and YMMV_Dsp sequences shared 96.3 % nt identity amongst themselves and 86.6-86.8%, respectively, with the YMMV sequence of *D. trifida* from Brazil. Bayesian analysis based on 67 nt sequences of the CP region revealed that YMMV_DA and YMMV_Dsp sequences clustered within a new proposed phylogroup IX, which includes isolates from Togo, Nigeria and Brazil. All sequences from this phylogroup are from *D. alata* and share 93-99% nt identities in the CP region. Last, ELISA results showed that YMMV from Brazil does not infect *D. cayennensis-rotundata*. Here we describe the first near-complete genomic sequence of YMMV of *D. alata* from Brazil. These results may shed light on the characterization of potyviruses that infect yams in Brazil.

1. Introduction

Yam (*Dioscorea* spp.) is an important economic crop mainly for smallholders and is widely cultivated, whose production can be hampered by viral diseases. At least 23 viral species infect yams (Aleman et al., 1996; Bömer et al., 2018, 2016; Eni, 2013; Filloux et al., 2015; Fuji

and Nakamae, 2000; Hayashi et al., 2017; Mambole et al., 2014; Marais et al., 2020; Menzel et al., 2014; Rabelo-Filho et al., 2013; Seal and Muller, 2007; Silva et al., 2019; Sukal et al., 2017; Umber et al., 2017; Zhang et al., 2016), and some can cause up to 65% of yield losses (Adeniji et al., 2012; Eni et al., 2010; Fuji and Nakamae, 2000; Kenyon et al., 2001).

Among the potyvirus species that infect yams are: *Yam mild mosaic virus* (YMMV) (Rabelo-Filho et al. 2013), *Yam mosaic virus* (YMV) (Aleman et al., 1996) *Chinese yam necrotic mosaic virus* (CYNMV) (Kondo and Fujita, 2012), *Japanese yam mosaic virus* (JYMV) (Fuji and Nakamae, 2000), *Yam chlorotic necrotic mosaic virus* (YCNV) (Lan et al. 2018; Zhang et al. 2016), *dioscorea alata macluravirus*, and *dioscorea esculenta macluravirus* (Filloux et al., 2015). *Yam mild mosaic virus* (YMMV) is a member of the *Potyvirus* genus, within the family *Potyviridae* (Adams et al., 2005; Wylie et al., 2017). Similar to other potyviruses, the genome of YMMV is a monopartite, single-stranded, positive-sense RNA molecule over 9.5 kb, with a viral genome-linked protein (VPg) covalently attached to its 5' end and a poly(A) tail at its 3' end (Fuji and Nakamae, 2000; Rabelo-Filho et al., 2013). The viral genome contains a large open reading frame (ORF) that is translated into a polypeptide which is cleaved into ten mature multifunctional proteins: protein 1 (P1), helper component-proteinase (HC-Pro), protein 3 (P3), six kilodalton 1 (6K1), cylindrical inclusion protein (CI), six kilodalton 2 (6K2), VPg, nuclear inclusion protein-a proteinase (NIa-Pro), nuclear inclusion b (NIb) and the coat protein (CP) (Adams et al., 2005; Wylie et al. 2017). Additionally, a conserved polymerase slip motif (5'-GAAAAAA-3') in the P3 cistron produces the P3N-PIPO protein (Chai et al., 2020; Olsper et al. 2015).

Mumford and Seal (1997) provided the first molecular evidence to classify YMMV as a distinct potyvirus (Mumford and Seal, 1997). In 1996 the complete genomic sequence of YMV was determined through RT-PCR (Aleman et al., 1996). Shortly thereafter, a new potyvirus infecting *D. alata* plants showing symptoms of mosaic in Papua New Guinea fields was detected by RT-PCR and identified as YMMV (Fuji et al. 1999; Mumford and Seal, 1997) by comparisons of the complete CP sequence and the 3'-UTR region (Fuji et al., 1999). The natural occurrence of YMMV in fields from China was first reported in 2011 (Zou et al. 2011). The presence of YMMV was reported in several yam-producing countries in Africa (Nkere et al. 2020; Odu et al., 1999), Asia and Oceania (Fuji et al. 1999; Mumford and Seal, 1997). By the 2000s, YMMV was identified in the Caribbean and South America (Bousalem et al., 2000; Dallot et al. 2001). In Brazil YMMV was first detected in *D. alata* and *D. trifida* (Bousalem et al. 2003; Pio-Ribeiro et al. 2006; Rabelo-Filho et al. 2013). However, only a partial genomic region of YMMV consisting of the C-terminal of NIb, complete CP and 3'-UTR (total 1056 nt)

(NCBI accession number: AF548498, isolate Br2) was amplified by IC-RT-PCR technique from *D. alata* collected in 2000 (Bousalem et al. 2003). Later, in 2009, YMMV was detected by RT-PCR from samples of *D. trifida* showing mosaic symptoms collected in the city of Paudalho, State of Pernambuco (Rabelo-Filho et al. 2013). The complete genomic sequence of the Brazilian isolate of YMMV was characterized in 2013 from *D. trifida* and consisted of 9538 nt (NCBI accession number: NC_019412, isolate Brazil).

In addition to infecting *D. alata* and *D. trifida*, YMMV has also been found to infect *D. rotundata* (Bousalem et al. 2003; Nkere et al. 2020), *D. opposita* (Zou et al. 2020), *D. esculenta* (Bousalem et al. 2003), and *D. polystachya* (Wang et al. 2015). The virus can be spread mainly through vegetative routes, such as tubers or cuttings from infected plants. Transmission of YMMV by aphids has already been reported (Odu et al. 1999), but the importance of insect vectors in transmission has not yet been established. YMMV, in addition to naturally infecting different species of *Dioscorea*, can also be mechanically transmitted to cowpea cultivars (Odu et al., 1999). *Dioscorea* species infected with YMMV may show mild mottling and mosaic symptoms or are asymptomatic (Kenyon et al., 2001; Mumford and Seal, 1997; Zou et al., 2020).

Bousalem et al. (2003) performed phylogenetic analysis using 36 sequences of 798 nt of CP and found eight clusters (clusters I - VIII). Later, Nkere et al. (2020) collected *dioscorea* samples from Guinea, Nigeria, Benin and Cameroon (from 2012 to 2013) to detect YMMV. They proposed additional phylogroups (IX, X, and XI) based on the complete sequence of CP. In this study, we report the identification and genetic characterization of two near-complete YMMV genomic sequences from Brazil assembled from HTS data. Furthermore, based on the comprehensive phylogenetic analysis of 67 CP complete nt sequences, we propose a restructuring of the YMMV groups into ten phylogroups (I-X). Finally, we showed that the Brazilian YMMV sequences clustered with other sequences from *D. alata* and are distantly related to sequences isolated from *D. trifida*. These results may shed light on the characterization of potyviruses that infect yams in Brazil.

2. Material and methods

2.1. Sample collection and ELISA survey

From October 2018 to January 2020, surveys were carried out to reveal infection of YMMV in yam fields in Brazil. A total of 75 young symptomatic and asymptomatic *D. alata* leaves (one leaf/plant) were collected as follows: Maranhão = 02, Paraíba = 39, and Pernambuco = 34. Samples were dried over calcium chloride (CaCl₂) and kept at 4 °C until further use for

the HTS sequencing. Seven samples of different plants of *D. alata* from the states of Pernambuco and Paraíba were grouped in a pool called "pool DA".

Double antibody sandwich (DAS)-ELISA was performed using the DSMZ ELISA kits (YMV, AS-1142; YMMV, AS-1038) following laboratory routines and recommendations associated with the assays. Positive controls for YMV (PC-1142) and YMMV (PC-1214) and negative controls were included in each ELISA test. Samples were considered positive when absorbance values were three times the mean of absorbance values of negative controls.

2.2. RNA extraction and High-Throughput Sequencing

Total RNA was extracted with the GenCatchTM Plant Total RNA Miniprep Kit (Epoch Life Sciences, USA) and RNA quality was evaluated by agarose gel electrophoresis. After ribosomal RNA depletion using the RiboMinusTM Plant Kit (ThermoFischer Scientific, Carlsbad, CA), cDNA and second-stranded synthesis were performed using random octamer primers. Libraries were constructed using a Nextera XT DNA library preparation kit (Illumina Inc., San Diego, CA) and subjected to MiSeq paired-end reads sequencing (2 x 301 bp, DSMZ, Germany).

An additional dataset generated by HiSeq sequencing (100 bp, paired-end, Illumina Inc. Macrogen, Seoul, Korea) of a mixed sample ("Dsp") of *D. cayennensis-rotundata* and *D. alata* plants from the states of Pernambuco and Paraíba from 2014 were also considered in this work.

2.3. Viral genome assembly and sequence analysis

The qualities of RNA-seq readings were checked with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were trimmed using Trimmomatic v.0.39 (Bolger et al. 2014) with the following parameters: ILLUMINACLIP:Nextera:2:30:10, LEADING:20, TRAILING:20, SLIDINGWINDOW:4:20, MINLEN:36. Trimmed reads were then assembled with SPAdes v.3.12.0 (Bankevich et al. 2012) using *K-mer* values of 21, 33, 55, and 77. Assembled contigs were subjected to tBlastX searches (E-value $1e^{-10}$, Matrix BLOSUM62, Word Size 3) against a manually assembled viral reference genome database (RefSeq March 2021, NCBI) implemented in Geneious v. R11 (Biomatters, Auckland, New Zealand). Contigs resulting in a hit after tBlastX searches with identities equal or higher than 70% to YMMV genomic sequences were selected and *de novo* assembled using the "De Novo Assemble" function (parameters: medium-low/fast sensitivity). The *de novo* assembled consensus was mapped with trimmed reads using the "map-to-reference" functions from Geneious: Medium-Low Sensitivity/Fast, fine-tuning of Iterate up to 5 times; Bowtie2 (High Sensitivity/Medium), and BBDMap (Normal Sensitivity). The assembled

YMMV full-length genomic sequence was assessed for putative ORFs using the NCBI ORF finder (<https://www.ncbi.nlm.nih.gov/orffinder/>), and for putative conserved domains with the Conserved Domain Database (CDD) (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Geneious software v. R11 (Biomatters, Auckland, New Zealand) was used to perform multiple sequence alignments using the implemented MAFFT v.7 algorithm (parameters: Algorithm: Auto, scoring matrix: 200PAM/k=2, Gao open penalty = 1.53, Offset value: 0.108). Potential cleavage sites in the YMMV polyprotein were identified from multiple sequence alignments using the following reference sequences from Brazil (NC_019412, Rabelo-Filho et al., 2013), China (KC407674-CN1, KC473517-CN20, KJ125478-WX3, KJ125479-XZ1, KJ125475-WX1, KJ125476-NC1, KJ125477-NC2, KJ125477-NC2, KJ125472-NN1, KJ125473-NC3, KJ125474-FX1, Wang et al. 2015; Zou et al. 2020), and from South Korea (KX156847-Andong, Unpublished) (Table 1).

The Sequence Demarcation Tool (SDTv.1.2) (Muhire et al. 2014) was used for pairwise comparison analyses at the nucleotide and amino acid levels with other sequences retrieved from GenBank.

2.4. Bayesian inferences

A “dataset_67_CP” composed of YMMV sequences from GenBank and the two assembled YMMV sequences from this work were used for Bayesian analyses. The “dataset_67_CP” (798 aligned sites) consisted of 65 CP nt sequences from GenBank and the two YMMV sequences from this work (Table 1).

The sequences from each dataset were aligned using the MAFFT v.7 algorithm (Katoh and Standley, 2013). The statistical selection of best-fit models of nucleotide substitutions was assessed with jModelTest v.2.1.10 (Darriba et al. 2012), indicating the Hasegawa–Kishino–Yano model (HKY) (Hasegawa et al. 1985) with a gamma distribution (G) of substitution rates as the best model. Phylogenetic trees were constructed using MrBayes v.3.2.7 (Ronquist et al. 2011), which implements a Markov chain Monte Carlo (MCMC) approach to approximate posterior probabilities of phylogenetic trees and clades. Each run was based on a chain of 5×10^6 generations, with a tree sampled every 100 generations (mcmc ngen = 5,000,000 samplefreq = 100). Plots of log likelihoods versus generation numbers were examined with Tracer v.1.7.1 (Rambaut et al. 2018) and the first 10 % of samples were determined as burn-in. Runs ceased when the standard deviation of split frequencies was below 0.004 and the Potential Scale Reduction Factor (PSRF) for the posterior probability distribution reached values of 1.0

166 (Ronquist et al. 2011). Phylogenetic trees were visualized and edited with FigTree v.1.4.3
167 (<http://tree.bio.ed.ac.uk/software/figtree/>) and Inkscape v. 1.0.2-2 (<https://www.inkscape.org/>).

168 **Table 1** YMMV sequences used for Bayesian inferences.

Code ¹	Isolate	Sequence Length (nt)	country	Collection date	host	Reference
MT345579_GHA	DaG4-43	798	Ghana	2012	<i>Dioscorea alata</i>	(Nkere et al., 2020)
MT345580_GHA	DaG89-11	798	Ghana	2012	<i>Dioscorea alata</i>	(Nkere et al., 2020)
MT345581_GHA	DaG131-16	798	Ghana	2012	<i>Dioscorea alata</i>	(Nkere et al., 2020)
MT345582_GHA	DaG137-17	798	Ghana	2012	<i>Dioscorea alata</i>	(Nkere et al., 2020)
MT345583_GHA	DrG269-31	798	Ghana	2012	<i>Dioscorea rotundata</i>	(Nkere et al., 2020)
MT345584_NIG	DaN31-3	798	Nigeria	2012	<i>Dioscorea alata</i>	(Nkere et al., 2020)
MT345585_NIG	DaN32-3	798	Nigeria	2012	<i>Dioscorea alata</i>	(Nkere et al., 2020)
MT345586_NIG	DaN143-12	798	Nigeria	2012	<i>Dioscorea alata</i>	(Nkere et al., 2020)
MT345587_NIG	DaN246-21	798	Nigeria	2012	<i>Dioscorea alata</i>	(Nkere et al., 2020)
MT345588_NIG	DaN7-19	798	Nigeria	2012	<i>Dioscorea alata</i>	(Nkere et al., 2020)
MT345589_NIG	DaN6-I	798	Nigeria	2012	<i>Dioscorea alata</i>	(Nkere et al., 2020)
MT345592_GHA	DaG4-43	798	Ghana	2013	<i>Dioscorea alata</i>	(Nkere et al., 2020)
MT345594_NIG	DaN46-5	798	Nigeria	2013	<i>Dioscorea alata</i>	(Nkere et al., 2020)
MT345595_NIG	DaN48-5	798	Nigeria	2013	<i>Dioscorea alata</i>	(Nkere et al., 2020)
MT501392_CAM	HB86	798	Cameroon	2014	<i>Dioscorea esculenta</i>	Unpublished
MT501393_CAM	HB148	798	Cameroon	2014	<i>Dioscorea alata</i>	Unpublished
KC924403_BRA	I-133	801	Brazil	2013	<i>Dioscorea alata</i>	Unpublished
AF548515_FIJ	FIJI 1	964	Fiji	2000	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548516_FIJ	FIJI 2	964	Fiji	2000	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548517_FIJ	FIJI 3	964	Fiji	2000	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548518_FIJ	FIJI 4	964	Fiji	2000	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548522_JAP	JPN 4	988	Japan	1993	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548525_PNG	PNG 3	1000	Papua New Guinea	1993	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548526_PNG	PNG4	1002	Papua New Guinea	1994	<i>Dioscorea alata</i>	(Bousalem et al. 2003)

AF548524_PNG	PNG2	1014	Papua New Guinea	1993	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548520_JAP	JPN2	1017	Japan	1998	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548521_JAP	JPN3	1019	Japan	1998	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548509_FGU	Guy4	1053	French Guiana	1998	<i>Dioscorea trifida</i>	(Bousalem et al. 2003)
AF548491_COL	Col1	1056	Colombia	1999	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548492_COL	Col2	1056	Colombia	1999	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548493_MAR	Mart1	1056	Martinique	1998	<i>Dioscorea rotundata</i>	(Bousalem et al. 2003)
AF548495_MAT	Mart3	1056	Martinique	1998	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548496_MAR	Mart4	1056	Martinique	1998	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548497_MAR	Mart5	1056	Martinique	1998	<i>Dioscorea rotundata</i>	(Bousalem et al. 2003)
AF548498_BRA	Br2	1056	Brazil	2000	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548500_GUA	Guad1	1056	Guadeloupe	1998	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548501_GUA	Guad2	1056	Guadeloupe	1998	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548502_GUA	Guad3	1056	Guadeloupe	1998	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548506_FGUI	Guy1	1056	French Guiana	1998	<i>Dioscorea trifida</i>	(Bousalem et al. 2003)
AF548510_FGUI	Guy5	1056	French Guiana	1998	<i>Dioscorea trifida</i>	(Bousalem et al. 2003)
AF548511_FGUI	Guy6	1056	French Guiana	1998	<i>Dioscorea trifida</i>	(Bousalem et al. 2003)
AF548513_FGUI	Guy8	1056	French Guiana	1998	<i>Dioscorea trifida</i>	(Bousalem et al. 2003)
AF548503_GUA	Guad4	1057	Guadeloupe	1998	<i>Dioscorea trifida</i>	(Bousalem et al. 2003)
AF548507_FGUI	Guy2	1057	French Guiana	1998	<i>Dioscorea trifida</i>	(Bousalem et al. 2003)
AF548508_FGUI	Guy3	1057	French Guiana	1998	<i>Dioscorea trifida</i>	(Bousalem et al. 2003)
AF548512_FGUI	Guy7	1057	French Guiana	1998	<i>Dioscorea trifida</i>	(Bousalem et al. 2003)
AF548514_TOG	Togo1	1057	Togo	1999	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548494_MAT	Mart2	1059	Martinique	1998	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548505_GUA	Guad6	1059	Guadeloupe	1998	<i>Dioscorea esculenta</i>	(Bousalem et al. 2003)
AF548519_JAP	JPN1	1059	Japan	1997	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548523_PNG	PNG1	1059	Papua New Guinea	1993	<i>Dioscorea alata</i>	(Bousalem et al. 2003)
AF548499_CRI	CR1	1060	Costa Rica	1999	<i>Dioscorea trifida</i>	(Bousalem et al. 2003)
JF357963_CHI	Luzhai	1278	China	2010	<i>Dioscorea opposita</i>	(Zou et al., 2020)

YMMV_Dsp	YMMV_Dsp	9501	Brazil	2014	<i>Dioscorea</i> sp.	This work
YMMV_DA	YMMV_DA	9512	Brazil	2019	<i>Dioscorea alata</i>	This work
KX156847_SKO	Andong	9521	South Korea	2014	<i>Dioscorea polystachya</i>	Unpublished
KC407674_CHI	CN1	9527	China	2011	<i>Dioscorea opposita</i>	(Wang et al. 2015)
KC473517_CHI	CN20	9529	China	2010	<i>Dioscorea polystachya</i>	(Wang et al. 2015)
KJ125478_CHI	WX3	9529	China	2012	<i>Dioscorea alata</i>	(Zou et al. 2020)
KJ125479_CHI	XZ1	9529	China	2012	<i>Dioscorea alata</i>	(Zou et al. 2020)
KJ125475_CHI	WX1	9530	China	2011	<i>Dioscorea opposita</i>	(Zou et al. 2020)
KJ125476_CHI	NC1	9533	China	2011	<i>Dioscorea alata</i>	(Zou et al. 2020)
KJ125477_CHI	NC2	9534	China	2011	<i>Dioscorea alata</i>	(Zou et al. 2020)
KJ125472_CHI	NN1	9536	China	2011	<i>Dioscorea alata</i>	(Zou et al. 2020)
KJ125473_CHI	NC3	9536	China	2011	<i>Dioscorea alata</i>	(Zou et al. 2020)
KJ125474_CHI	FX1	9537	China	2012	<i>Dioscorea</i> sp.	(Zou et al. 2020)
NC_019412_BRA	Brazil	9538	Brazil	2009	<i>Dioscorea trifida</i>	(Filho et al. 2013)

¹Code: NCBI Accession_Country code. BRA: Brazil, CAM: Cameroon, CHI: China, COL: Colombia, CRI:Costa Rica, FGUI: French Guiana, FIJ: Fiji, GHA: Ghana, GUA: Guadeloupe, JAP: Japan, MAR: Martinique, NIG: Nigeria, PNG: Papua New Guinea, SKO: South Korea, TOG: Togo.

The bold letters are sequences from Brazil.

2.5. Genomic sequence recombination analysis

First, the “dataset_14_complete” was aligned using the MAFFT v.7 algorithm (Kato and Standley, 2013). After, the Recombination Detection Program version 4 (RDP4) with default settings for the different detection methods and a Bonferroni corrected *P*-value cut-off of 0.05 (Martin et al. 2015) and seven detection methods including RDP, GENECONV (Padidam et al. 1999), BootScan (Martin et al. 2005), MaxChi (Smith, 1992), Chimaera (Posada and Crandall, 2001), SiScan (Gibbs et al. 2000), and 3Seq (Lam et al. 2018) were applied to detect putative recombination breakpoints among the 14 sequences.

3. Results and Discussion

3.1. Symptoms and ELISA results

In all inspected cultivated *D. alata* fields vigorous plants without clear symptoms on leaves were observed. Few plants showed weak mosaic symptoms that were barely visible on leaves (Fig. 1).

According to ELISA results, 64.1% (25/39) of the collected leaves from Paraíba and 82.4% (28/34) from Pernambuco (data not shown) were infected with YMMV. No YMMV was detected in samples from Maranhão. According to the serological tests YMMV was detected only in *D. alata* plants. In contrast, YMV were not detected in samples of *D. alata* (pool DA).

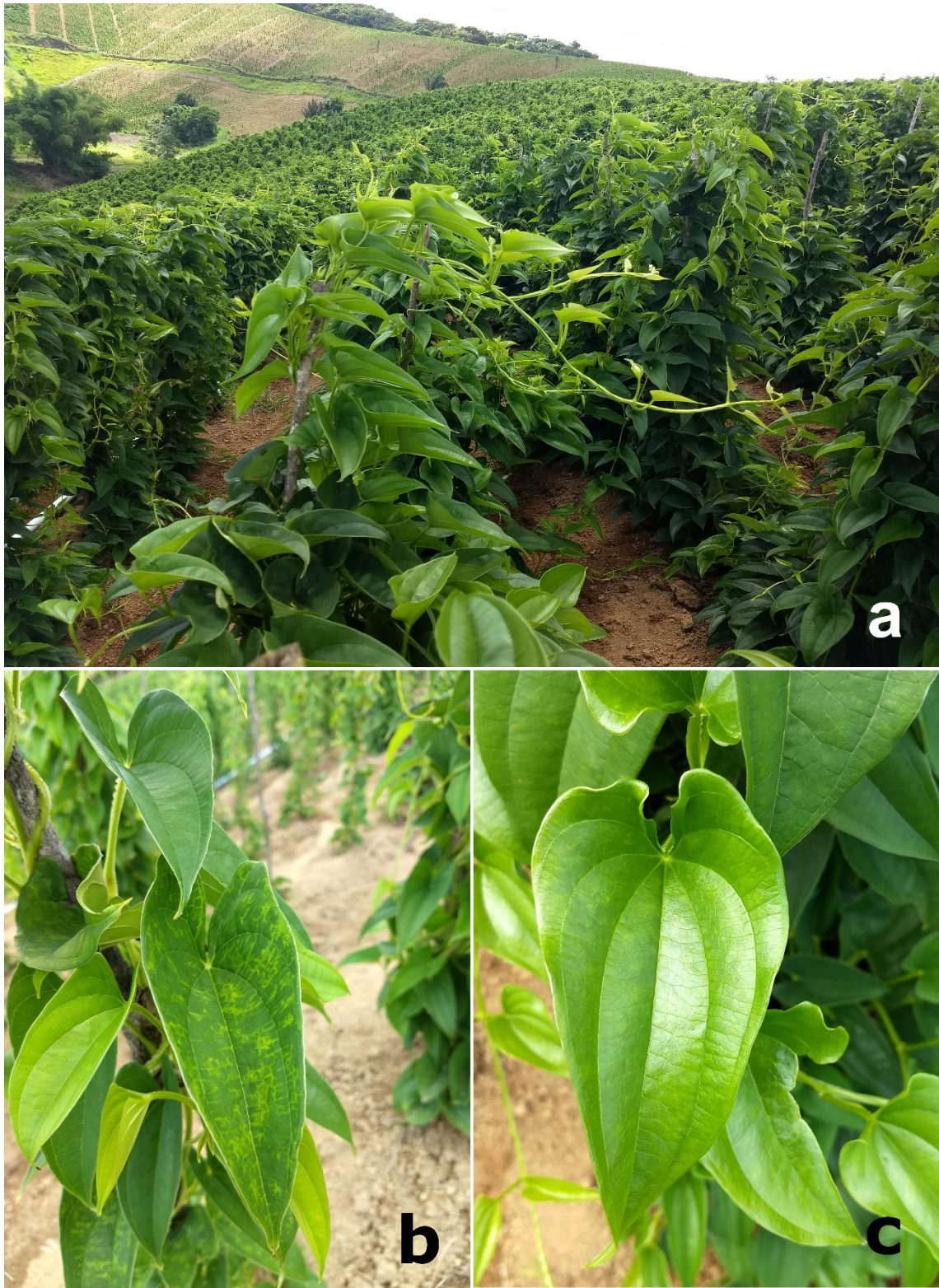


Fig. 1 *Dioscorea alata* from yam growing areas in Brazil. **a.** field from Bonito, Pernambuco **b.** symptomatic leaves, and **c.** asymptomatic leaves.

3.2. Characterization of YMMV genomic sequences

Two genomic sequences of YMMV were assembled from each sequenced pool (DA, and Dsp) assigned as YMMV_DA and YMMV_Dsp, respectively. YMMV_DA genomic sequence comprises 9,512 nt in length, while YMMV_Dsp 9,501 nt, excluding poly-(A) tails (Table 2). The largest ORF of all assembled YMMV genomic sequences consisted of 9,255 nt that putatively encodes a polyprotein of 3,084 amino acids (aa) with an estimated molecular weight of 342,32 kDa.

The predicted nine polyprotein cleavage sites were obtained through multiple alignments of YMMV genomic sequences available at NCBI based on the proposed criteria presented by Adams et al. (2005). The predicted cleavage sites in the YMMV polyprotein are depicted in Figure 2. A second small ORF (P3N-PIPO) was predicted from an A insertion between nt positions 2,910 and 2,915 (GA₂₉₁₀A_AAA_A₂₉₁₅TT) of the YMMV_DA, and 2,911 and 2,916 (GA₂₉₁₁A_AAA_A₂₉₁₆TT) of the YMMV_Dsp genomic sequences, respectively, resulting in a putatively expressed transcript of 204 nt long (Table 2).

The YMMV_DA and YMMV_Dsp polyprotein cleavage sites were predicted from multiple genomic nt sequence alignments of YMMV, and were as follows: ITHY₃₂₀/S (P1/HC-Pro); YRVG₇₇₇/G (HC-Pro/P3); VEHQ₁₁₂₅/A (P3/6K1); VVHQ₁₁₇₆/S (6K1/CI); VLHQ₁₈₁₉/S (CI/6K2); VTHQ₁₈₇₁/G (6K2/VPg); VIHE₂₀₆₀/G (VPg/NlA-Pro); VVEQ₂₃₀₂/S (NlA-Pro/NlB); VVHQ₂₈₁₈/A (NlB/CP). Interesting, the cleavage sites of 6K1/CI and NlB/CP differed from those described for YMMV (NC_019412) of *D. trifida* (VAHQ₁₁₇₆/S for 6K1/CI and VIHQ₂₈₁₈/A for NlB/CP) (Rabelo-Filho et al. 2013) from Brazil. The other putative seven cleavage sites showed the same amino acid sequence cleavage pattern.

All YMMV genomic sequences employed in this work showed the same length for the largest ORF (9255 nt in length), with exception of the one from South Korea (KC473517, 9258 nt). In all cases (dataset_14_complete), there was a slight difference in length for the untranslated regions (UTRs).

Table 2 Genomic characterization, putative proteolytic cleavage sites, and conserved motifs of the YMMV sequences generated in this study.

Name	Type	YMMV_DA			YMMV_Dsp		
		Initial position	Final position	Length	Initial position	Final position	Length
5'UTR	5'UTR	1	115	115	1	116	116
Polyprotein	CDS	116	9370	9255	117	9371	9255
P1	CDS	116	1075	960	117	1076	960
P1/HC_Pro: ITHY ₃₂₀ /S	Cleavage	1064	1078	15	1065	1079	15
HC-Pro	CDS	1076	2446	1371	1077	2447	1371
RITC ₃₇₄	motif	1226	1237	12	1227	1238	12
FRNK ₅₀₃	motif	1613	1624	12	1614	1625	12
CCC ₆₁₆	motif	1946	1954	9	1947	1955	9
PTR ₆₃₁	motif	2000	2008	9	2001	2009	9
HC_Pro/P3: YRVG ₇₇₇ /G	Cleavage	2435	2449	15	2436	2450	15
P3	CDS	2447	3490	1044	2448	3491	1044
PIPO	CDS	2911	3114	204	2912	3115	204
P3/6K1: VEHQ ₁₁₂₅ /A	Cleavage	3479	3493	15	3480	3494	15
6K1	CDS	3491	3643	153	3492	3644	153
6k1/CI: VVHQ ₁₁₇₆ /S	Cleavage	3632	3646	15	3633	3647	15
CI	CDS	3644	5572	1929	3645	5573	1929
GSGKS-X ₃ -P ₁₂₇₁	motif	3902	3928	27	3903	3929	27
DECH ₁₃₅₂	motif	4160	4171	12	4161	4172	12
SATPP ₁₃₈₂	motif	4247	4261	15	4248	4262	15
VATNIENGVTL ₁₄₈₉	motif	4547	4582	36	4548	4583	36
GERIQLGRVGR ₁₅₃₃	motif	4680	4714	35	4680	4715	36
CI/6K2: VLHQ ₁₈₁₉ /S	Cleavage	5561	5575	15	5562	5576	15
6K2	CDS	5573	5728	156	5574	5729	156
6K2/VPg: VTHQ ₁₈₇₁ /G	Cleavage	5717	5731	15	5718	5732	15
VPg	CDS	5729	6295	567	5730	6296	567
VPg/Nla-Pro: VIHE ₂₀₆₀ /G	Cleavage	6284	6298	15	6285	6299	15
Nla-Pro	CDS	6296	7021	726	6297	7022	726
Nla-Pro/Nlb: VVEQ ₂₃₀₂ /S	Cleavage	7010	7024	15	7011	7025	15
Nlb	CDS	7022	8569	1548	7023	8570	1548
FTAAPID ₂₄₉₆	motif	7583	7603	21	7584	7604	21
CVDDFN ₂₅₀₉	motif	7625	7642	18	7626	7643	18
GNNSGQP ₂₆₁₄	motif	7937	7957	21	7938	7958	21
GDD ₂₆₅₃	motif	8066	8074	9	8067	8075	9
Nlb/CP: VVHQ ₂₈₁₈ /A	Cleavage	8558	8572	15	8559	8573	15
CP	CDS	8570	9370	801	8571	9371	801
DAG ₂₈₂₈	motif	8591	8599	9	8592	8600	9
3'UTR	3'UTR	9371	9512	142	9372	9501	130

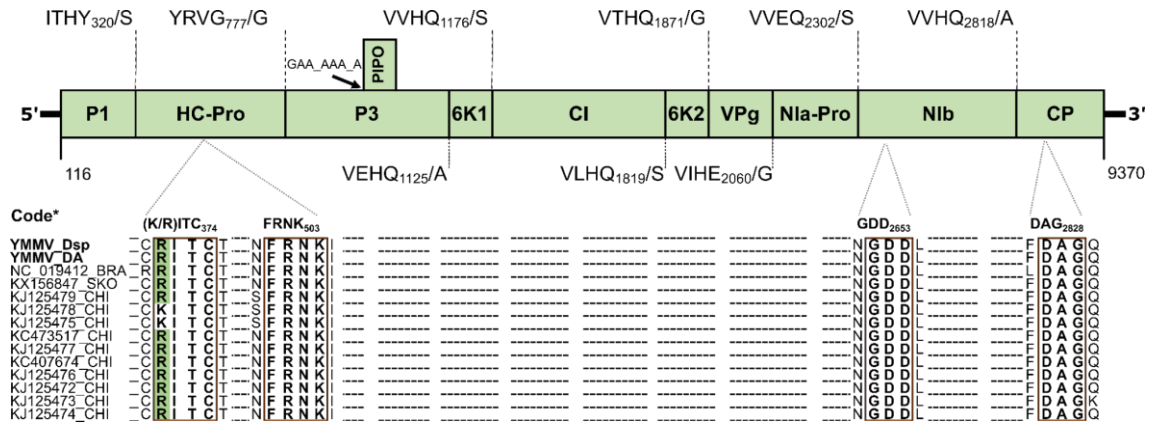


Fig. 2 Schematic representation of YMMV genome organization, predicted cleavage sites and known conserved motifs among 14 aligned genomic sequences. Potyvirus motifs: (K/R)ITC, FRNK in HC-Pro; GDD in Nlb; and DAG in the CP.

Code* bold letters represent sequences of this work. Not bold are represented by NCBI Accession_Country code. BRA: Brazil, CHI: China, and SKO: South Korea.

The following known conserved potyvirus motifs (Lan et al. n.d.; Revers and García, 2015; Shibolet et al. 2007; Torrance et al. 2006; Worrall et al. 2019; Wylie et al. 2017) were identified in the YMMV polyprotein sequences of this study: FRNK₅₀₃ (involved in RNA silencing suppression), and CCC₆₁₆ (long-distance movement) in HC-Pro; GSGKS-X3-P₁₂₇₁, DECH₁₃₅₂, SATPP₁₃₈₂, VATNIIENGVTL₁₄₈₉, GERIQLRGRVGR₁₅₃₃ (potential helicase activity) and SATPP₁₃₈₂ in CI; FTAAPID₂₄₉₆, CVDDFN₂₅₀₉, GNNSGQP₂₆₁₄, and GDD₂₆₅₃ (RNA-dependent polymerase activity) in Nlb and the DAG₂₈₂₈ (aphid transmission) motif in CP (Table 3). Interestingly, the conserved KITC (Lysine/Isoleucine/Threonine/Cysteine) motif located at the N-terminal domain of HC-Pro, which is involved in specific binding to receptors located in the extreme tip of the stylet called the acrostyle (Blanc et al. 1998; Uzest et al. 2007), were not found in YMMV sequences (Table 3). Instead, a mutation at the second nt (A-G) of the code changed this domain to RITC₃₇₂ (Fig. 2 and Table 3). This motif was already described for the Brazilian YMMV sequence (NC_019412) of *D. trifida* (Rabelo-Filho et al. 2013), and other potyviruses such as tobacco vein banding mosaic virus (TVBMV, isolate Yunnan-China) (Yu et al. 2007), bean yellow mosaic virus (BYMV) (Wylie et al. 2002), pecan mosaic associated virus (PMaV from *Carya illinoensis*) (Su et al. 2016), and plum pox virus (isolate Y21 from Japan) (Karavina et al. 2020).

256 **Table 3** Conserved motifs of YMMV sequences analyzed from South Korea, China, and Brazil.

Sequences	(K/R)ITC ₃₇₄	FRNK ₅₀₃	PTR ₆₃₁	SATPP ₁₃₈₂	VATNIENGVTL ₁₄₈₉
YMMV_Dsp	C R I T C T	N F R N K I	I T P T R N	V S A T P P G I	V A T N I I E N G V T L D
YMMV_DA	C R I T C T	N F R N K I	I T P T R N	V S A T P P G I	V A T N I I E N G V T L D
NC_019412_BRA	R R I T C T	N F R N K I	I T P T R N	V S A T P P G I	V A T N I I E N G V T L D
KX156847_SKO	C R I T C T	N F R N K I	I T P T R N	V S A T P P G I	V A T N I I E N G V T L D
KJ125479_CHI	C R I T C T	S F R N K I	I T P T R N	V S A T P P G V	V A T N I I E N G V T L D
KJ125478_CHI	C K I T C T	S F R N K I	I T P T R N	V S A T P P G V	V A T N I I E N G V T L D
KJ125475_CHI	C K I T C T	S F R N K I	I T P T R N	V S A T P P G V	V A T N I I E N G V T L D
KC473517_CHI	C R I T C T	N F R N K I	I T P T R N	V S A T P P G V	V A T N I I E N G V T L D
KJ125477_CHI	C R I T C T	N F R N K I	I T P T R N	V S A T P P G V	V A T N I I E N G V T L D
KC407674_CHI	C R I T C T	N F R N K I	I T P T R N	V S A T P P G V	V A T N I I E N G V T L D
KJ125476_CHI	C R I T C T	N F R N K I	I T P T R N	V S A T P P G I	V A T N I I E N G V T L D
KJ125472_CHI	C R I T C T	N F R N K I	I T P T R N	V S A T P P G I	V A T N I I E N G V T L D
KJ125473_CHI	C R I T C T	N F R N K I	I T P T R N	V S A T P P G I	V A T N I I E N G V T L D
KJ125474_CHI	C R I T C T	N F R N K I	I T P T R N	V S A T P P G I	V A T N I I E N G V T L D

Sequences	QRLGRVGR ₁₅₃₃	GNNSGQP ₂₆₁₄	GDD ₂₆₅₃	DAG ₂₈₂₈
YMMV_Dsp	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	F D A G Q
YMMV_DA	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	F D A G Q
NC_019412_BRA	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	L D A G Q
KX156847_SKO	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	F D A G Q
KJ125479_CHI	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	F D A G Q
KJ125478_CHI	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	F D A G Q
KJ125475_CHI	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	F D A G Q
KC473517_CHI	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	F D A G Q
KJ125477_CHI	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	F D A G Q
KC407674_CHI	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	F D A G Q
KJ125476_CHI	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	F D A G Q
KJ125472_CHI	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	F D A G Q
KJ125473_CHI	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	F D A G K
KJ125474_CHI	I Q R L G R V G R H	R G N N S G Q P S	N G D D L	F D A G Q

257

258 The bold letters are sequences of this work (YMMV_DA, YMMV_Dsp). Not bold (NCBI Accession_Country code), BRA: Brazil, CHI: China, and SKO: South Korea.

All CP YMMV sequences contained the DAG₂₈₂₈ motif, which is known to be important for aphid transmission (Atreya et al. 1995; Fuji et al. 1999). This motif interacts directly with the PTK (Phe-Thr-Lys) motif located at the C-terminus of HC-Pro (Huet et al., 1994; Peng et al. 1998) to facilitate the binding of HC-Pro to the coat protein. Interestingly, the PTK motif in HC-Pro that should be located at the C-terminus of HC-Pro (Peng et al. 1998), was not found in the analyzed YMMV sequences. Instead, a PTR₆₃₁ motif was found. This PTR motif was also found in the potyvirus sequence of sudan watermelon mosaic virus (SuWMV, isolate Su94-54, NC_035459) (Desbiez et al. 2017) from samples collected in 1994 from snake cucumber (*Cucumis melo* var. *flexuosus*) in Eastern Sudan near Gadamballia.

3.3. Pairwise sequence analyses of the Brazilian genomes

SDT analysis of YMMV_DA and YMMV_Dsp genomic sequences showed 86.6 and 86.8% nt identity with the complete YMMV sequence of *D. trifida* from Brazil (NC_019412), and 86 and 86.3 % nt identities with the sequence from South Korea (Andong), respectively. The Brazilian YMMV sequences shared 83.6 to 85.6% nt identities with sequences from China (CN1, CN20, NN1, NC3, FX1, WX1, NC1, NC2, WX3, and XZ1). They are genetically more distant from sequences from South China (CN1, NC2, NN1, NC1, NC3, FX1 and NC20) as they share only 83.7-85% nt identities, followed by sequences from North China (WX1, WX3 and XZ1) sharing nt identities of 85.3-85.6%. Comparisons of nucleotide and amino acid sequences of coding regions of YMMV_DA and YMMV_Dsp with chinese sequences resulted in the lowest values of identities of 78-80% at the nt level and 74-82% at aa level for the P1 region (Table 5).

Finally, pairwise comparison results of 14 YMMV genomic sequences ("dataset_14_complete", 9475 aligned sites) show that the YMMV_DA and YMMV_Dsp sequences were closely related to sequences from Brazil (NC_019412) and South Korea (Andong) (86-86.8% nt identities). Detailed pairwise sequence comparisons among complete genomic sequences are shown in Table 5.

Table 5 Pairwise comparisons of nucleotides and deduced amino acid sequences of genomic sequences, polyproteins, individual coding and non-coding regions of YMMV from this study (YMMV_DA, and YMMV_Dsp) and reference sequences using SDT software.

Isolates	P1		HC-Pro		P3		PIPO		6K1		CI		6K2		NIa		VPg		NIb		CP		POLY		5'UTR	3'UTR	WG
	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	nt	nt
YMMV_DA																											
YMMV_Dsp	98.6	97.5	98.5	99.8	96.2	99.4	100	100	99.3	100	95.1	98.4	98.1	100	96.1	97.9	98.6	100	95.2	98.1	93.3	94.4	96.4	98.3	98.3	91.5	96.3
NC_019412_BRA (Brazil)	98.4	97.5	87.4	96.7	84.7	91.4	95.1	92.5	84.3	96.1	84.9	94.4	85.3	90.4	84.2	9.5	84.5	95.8	84.8	93.8	84.0	91.0	86.5	94.4	95.7	89.4	86.6
KC407674_CHI (CN1)	79.1	80.3	86.2	97.2	82.7	87.4	90.2	82.1	79.1	96.1	86.6	96.6	85.9	94.2	86.5	95.0	84.5	95.2	85.1	94.2	85.3	95.9	84.7	93.2	80.0	83.8	84.6
KC473517_CHI (CN20)	79.4	74.1	86.7	98.0	84.6	91.4	91.7	89.6	81.7	94.1	86.8	96.6	85.3	96.2	85.3	96.7	84.3	95.2	84.4	94.8	87.8	94.7	85.1	93.3	79.1	82.4	84.9
KJ125472_CHI (NN1)	77.7	76.3	86.4	98.0	83.6	87.9	90.2	86.6	86.9	98.0	85.2	96.4	85.9	94.2	86.0	96.7	83.4	95.2	84.1	93.6	85.0	94.7	84.2	92.9	84.3	85.2	84.2
KJ125473_CHI (NC3)	78.2	76.6	85.0	97.4	84.3	88.2	91.2	86.6	85.0	98.0	85.6	96.9	84.0	94.2	86.1	95.9	85.7	95.8	84.8	93.4	85.9	95.5	84.5	93.0	80.0	86.6	84.4
KJ125474_CHI (FX1)	79.1	79.7	84.0	97.4	81.8	86.2	89.2	83.6	82.4	98.0	85.3	95.6	85.3	94.2	84.7	95.9	85.7	96.8	83.6	92.1	85.5	95.1	83.7	92.6	78.9	90.8	83.7
KJ125475_CHI (WX1)	79.8	78.4	87.6	97.6	8.5	89.7	90.2	85.1	84.3	96.1	86.8	97.2	87.2	96.2	86.2	95.9	84.5	96.3	85.5	95.2	84.9	94.4	85.4	93.7	80.7	88.7	85.3
KJ125476_CHI (NC1)	77.9	76.9	86.6	98.0	84.0	87.6	90.7	86.6	86.3	96.1	85.6	96.3	85.9	94.2	84.7	96.3	82.9	95.2	84.0	93.8	85.3	95.5	84.3	93.0	82.6	85.8	84.2
KJ125477_CHI (NC2)	79.3	81.9	84.7	97.2	81.5	84.2	90.2	83.6	83.7	96.1	85.8	96.4	80.8	92.3	84.6	95.9	83.8	94.7	84.4	95.0	83.4	94.4	83.7	93.0	80.9	83.8	83.6
KJ125478_CHI (WX3)	79.6	77.8	87.6	1.0	85.0	89.7	90.2	85.1	84.3	96.1	86.8	97.2	87.2	96.2	86.1	95.9	84.5	96.3	85.4	95.0	84.9	94.4	85.3	93.6	73.5	89.4	85.3
KJ125479_CHI (XZ1)	80.2	78.4	86.8	97.4	85.4	89.9	90.2	85.1	83.7	96.1	87.2	97.0	89.1	98.1	86.2	95.5	84.7	95.8	84.9	95.2	84.8	94.4	85.4	93.6	72.6	89.4	85.3
KX156847_SKO (Andong)	83.1	82.8	89.9	98.2	84.4	92.5	91.7	91.0	82.4	98.0	85.9	97.2	87.8	96.2	85.4	96.7	85.4	95.2	86.2	95.2	86.8	95.1	86.1	94.6	86.1	87.3	86.0
YMMV_Dsp																											
NC_019412_BRA (Brazil)	97.7	96.3	86.7	96.5	85.5	91.7	95	93	85.0	96	85.3	94.9	85.3	90	85.3	95.0	84.5	96	84.9	94.0	84.1	91.4	86.6	94.5	95.7	92.3	86.8
KC407674_CHI (CN1)	79.4	81.6	86.4	97.4	83.8	87.6	90.2	82.1	78.4	96.1	86.7	96.9	85.3	94.2	86.5	95.0	85.2	95.2	85.5	94.8	84.4	95.5	84.9	93.5	81.9	86.2	84.9
KC473517_CHI (CN20)	79.8	75.6	86.2	97.8	85.8	91.7	91.7	89.6	81.0	94.1	87.1	96.9	84.6	96.2	85.4	95.5	83.6	95.2	84.6	94.8	86.8	93.6	85.1	93.3	79.3	81.5	85.0
KJ125472_CHI (NN1)	78.0	77.5	86.1	97.8	83.3	88.2	90.2	86.6	86.3	98.0	8.5	96.7	84.0	94.2	85.8	95.5	83.1	95.2	84.5	93.6	84.5	94.4	84.0	93.0	85.3	83.8	84.0
KJ125473_CHI (NC3)	78.3	77.2	84.7	97.6	84.4	88.5	91.2	86.6	84.3	98.0	85.5	97.2	83.3	94.2	86.6	94.6	85.4	95.8	85.1	93.4	86.5	94.4	84.5	93.0	81.0	85.4	84.5
KJ125474_CHI (FX1)	79.0	79.7	83.9	97.6	83.0	86.5	89.2	83.6	83.0	98.0	85.4	96.0	83.3	94.2	85.3	94.2	85.4	96.8	84.2	92.1	85.9	95.1	84.0	92.6	78.3	87.7	84.0
KJ125475_CHI (WX1)	79.8	78.8	87.3	97.4	85.9	89.9	90.2	85.1	83.7	96.1	87.1	97.7	86.5	96.2	86.8	94.6	84.5	96.3	85.7	95.7	85.3	94.4	85.6	93.8	80.9	90.8	85.6
KJ125476_CHI (NC1)	79.9	78.1	86.4	97.8	83.5	87.9	90.7	86.6	85.6	96.1	85.3	96.6	84.0	94.2	85.1	95.0	82.5	95.2	84.8	93.8	84.8	95.1	84.2	93.0	83.6	83.1	84.1
KJ125477_CHI (NC2)	79.2	82.2	84.9	97.4	82.7	84.5	90.2	83.6	83.0	96.1	86.2	96.4	8.0	92.3	84.8	94.6	83.4	94.7	84.9	95.0	84.5	93.6	84.1	92.9	81.9	83.1	84.0
KJ125478_CHI (WX3)	79.7	78.1	87.3	97.4	86.0	89.9	90.2	85.1	83.7	96.1	87.1	97.7	86.5	96.2	86.6	9.5	84.5	96.3	85.7	95.9	85.3	94.4	85.6	93.8	72.8	90.8	85.6
KJ125479_CHI (XZ1)	80.3	78.4	86.7	97.2	86.1	90.2	90.2	85.1	83.0	96.1	8.7	97.5	88.5	98.1	86.8	94.2	84.7	95.8	85.7	95.5	85.5	94.4	85.7	93.7	71.9	90.0	85.6
KX156847_SKO (Andong)	83.5	83.8	89.7	98.0	8.6	92.8	91.7	91.0	83.0	98.0	86.7	97.8	87.2	96.2	85.5	95.0	84.8	95.2	85.9	95.0	86.0	94.7	86.3	94.7	87.9	85.4	86.3

287

288 **Bold.** Sequences of this study.

3.4. Bayesian analyses

Based on the Bayesian analysis of “dataset_67_CP” (Table 1), ten phylogenetic groups for YMMV (I-X) are proposed (Fig. 3) and are supported by high posterior probability values. Group I is represented by an individual taxon (isolate Guad1, Guadeloupe country). Groups II (Cameron) and III (Nigeria and Ghana) are composed of isolates from Africa. Based on intra-group comparisons of pairwise nt sequences of the CP region these sequences shared 99% and 87% nt identity, respectively, among themselves. Group IV (89-98% CP nt identity) is composed of isolates from South America (French Guiana and Brazil), Central America (Costa Rica), and Oceania (Papua New Guinea). Group V (84-89% CP nt identity) is constituted of two isolates from the Caribbean (Guadeloupe). Group VI (90-100%) is formed by sequences of northern China and group VII (89-99%) from South China, while Group VIII (87-99%) with sequences from countries of East Asia and Oceania (Japan, South Korea and Papua New Guinea). Group IX (93-99%) is composed of sequences from this study (YMMV_DA, YMMV_Dsp), two isolates from Brazil (I-133 and Br2), and two sequences from West Africa (Togo and Nigeria). Finally, group X (89-99%) is the largest group and has isolates from several regions of the world such as East Asia (Japan), Oceania (Fiji and Papua New Guinea), South America (Colombia), Caribbean (Guadeloupe, Martinique) and Africa (Nigeria and Ghana).

In 2003, Bousalem et al. (2003) analyzed the CP gene of 36 isolates of YMMV, and concluded that due to the polytomy structure of the tree it was difficult to establish a relationship between the different clades. The authors showed individual taxa with no significant interrelationship support either with bootstrap (BP) and reliability percentages (PR) less than 50%. However, they proposed two large groups with high support values of BP and RP: a group named “cosmopolitans-VIII” that we assigned as group IX and X, and another second large group named “VII” (America and the Caribbean) that in this study we grouped as IV. Bousalem et al. suggested that by their second significant monophyletic group (Guy1, Guy2, Guy3, Guy4, Guy5, Guy6, Guy7, Guy8, CR1) (group “VII”) there was a clear host adaptation to *D. trifida*. However, the authors could not conclude the existence of a geographical specialization of group VII due to the overrepresentation of isolates of *D. trifida* from French Guiana.

In this work of the ten proposed phylogroups (I-X), some showed association with the geographic distribution (Fig. 3). For instance, group II and group III are composed of sequences from Africa. Isolates from group IV are not only geographically related, but also host related, as the sequences from the Americas are from *D. trifida*, which also includes an isolate from Brazil (NC_019412). Group V (Caribbean) also has isolates from *D. trifida* and *D. alata*. Other

groups that presented isolates with a closer genetic relationship with their geographic locations are group VI (all from northern China) and group VII (all from southern China). These results are in agreement with Zou et al. (2020) that performed a maximum likelihood phylogenetic analysis using 89 nt YMMV sequences of partial CP regions and 3'-UTRs (259–262 nucleotides). Fourteen groups were associated with their geographical distribution. Among them, a group of isolates from southern China (CN20, NC2, FX1, NC3, NN1, NC1, Luzhai; sequences of group VII from this study) and another group from northern China (WX1, WX3, XZ1; sequences of group VI from this study). Isolates of group VIII also demonstrated a relationship to the geographical sampling, since the isolates are from countries located in western Asia (Japan), and Oceania (Papua New Guinea). On the contrary, no association with their geographic location but with the host was found within group IX (includes YMMV_DA and YMMV_Dsp), since isolates are all from *D. alata* according to our ELISA results. Interestingly, the Brazilian YMMV sequence of *D. trifida* (NC_019412), despite also found in the northeast region of Brazil, is phylogenetically distant to group IX sharing a low nt sequence identity of 84% in the CP region (Table 5). Likewise, the group X, which is composed of isolates of several locations around the world, are obviously not geographically related. Also, in 2020, Nkere et al. (2020) performed a phylogenetic analysis based on the complete CP nt sequence of 18 YMMV isolates from Nigeria and Ghana. Eight sequences from Nigeria and four from Ghana clustered into a cosmopolitan phylogroup (assigned by Busalem et al. 2003 as group “VIII”, and as groups IX and X in this study). Therefore, Nkere et al. concluded that there was no correlation with the geographic location of these isolates.

In most groups of this work, there was no clear relationship in the specialization of the isolates of YMMV with the host. Finally, the sequences of this study YMMV_DA (*D. alata*), YMMV_Dsp (presumably of *D. alata* since we did not find YMMV in *D. caynnensis-rotundata* by ELISA) cluster with other isolates of *D. alata* from Brazil, Nigeria and Togo within group IX, and therefore, it is also tempting to speculate a specialization of YMMV to *D. alata*. Further studies will be necessary to understand the relationship of YMMV to *D. alata* from the northeast region of Brazil.

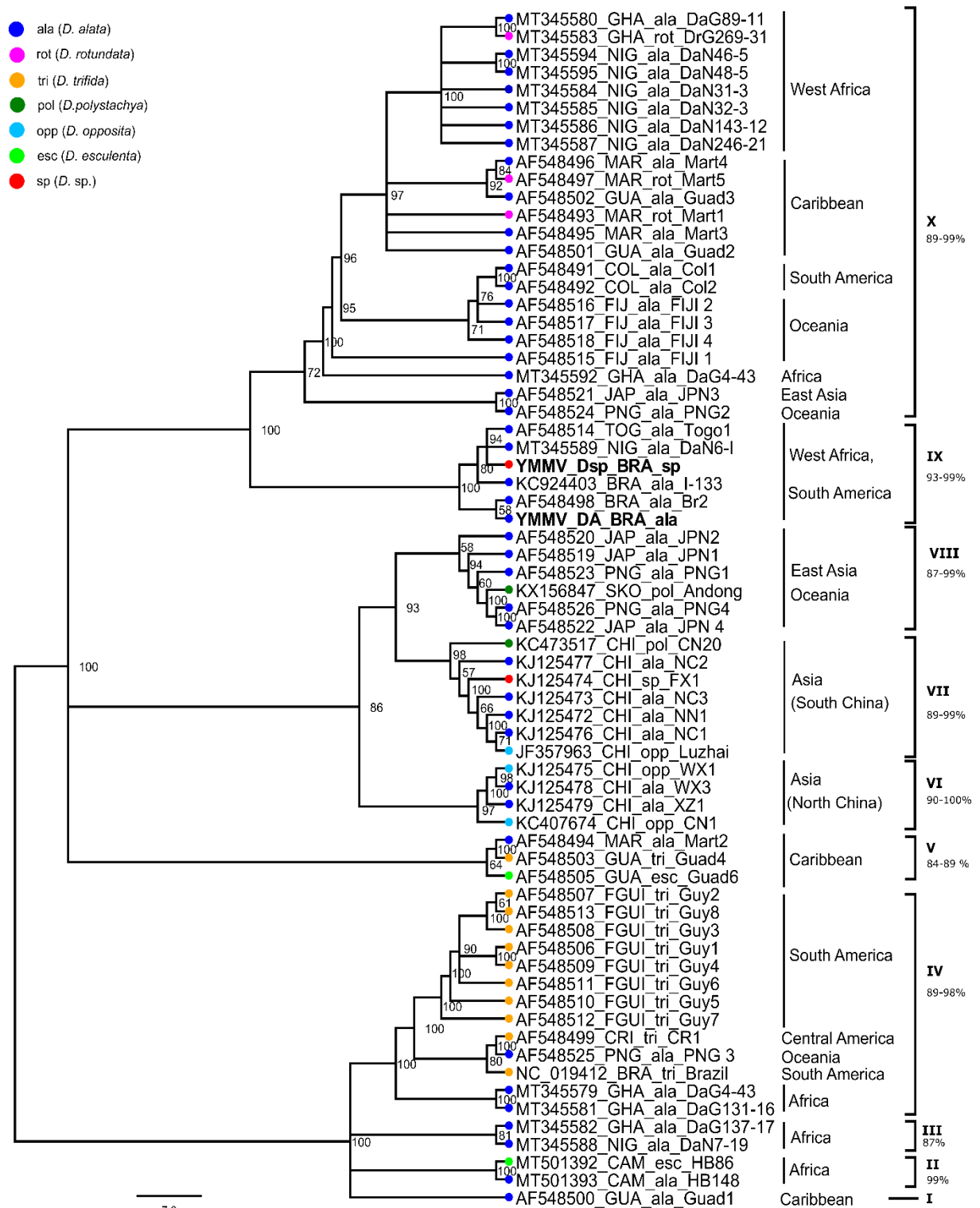


Fig. 3 Bayesian phylogenetic analysis of 67 complete CP nucleotide sequences of YMMV. Numbers at nodes represent mean posterior probability values, and the scale bar represents the nucleotide substitutions per site. Tip labels in bold correspond to the sequences of this study, those not in bold are represented by the following: GenBank acc._country_*Dioscorea* species_YMMV isolate. Species of *Dioscorea* are represented as *D. alata*=ala (blue), *D.*

esculenta = esc (light green), *D. opposita* = opp (light blue), *D. polystachya* = pol (dark green), *D. rotundata* = rot (pink), *D. trifida* = tri (orange), and *D. sp* = sp (red). Countries are BRA: Brazil, CAM: Cameroon, CHI: China, COL: Colombia, CRI: Costa Rica, FGUI: French Guiana, FIJ: Fiji, GHA: Ghana, GUA: Guadeloupe, JAP: Japan, MAR: Martinique, NIG: Nigeria, PNG: Papua New Guinea, SKO: South Korea, TOG: Togo. Ten numbers (I-X) represent the phylogroups. The percentages at the right-side of the tree report the pairwise identities inferred with the SDT v.1.2 software.

3.5. Genomic recombination events in YMMV

Alignment of fourteen YMMV genomic sequences "dataset_14_complete" did not reveal significant putative signals of recombination events (events 1, 3, and 5) (Table 6). Although at least five out of seven methods implemented in the RDP4 identified putative recombinant events with P -value less than 1.0×10^{-6} (Table 6), we did not consider these values as significant. In 2007, Ohshima et al. (2017) observed that 74% of 92 TuMV genomic sequences showed evidence of recombination in various regions of the genome. Found recombination sites between lineages in the P1, HC-Pro and P3 regions, which were supported with P -values of 16×10^{-9} - 16×10^{-19} (P1), 3.8×10^{-9} - 2.2×10^{-37} (HC-Pro), 5.0×10^{-30} - 1.1×10^{-36} (P3) with at least four methods using the RDP2 program. In another investigation, Braidwood et al. (2019) analyzed coding regions of 33 genomic sequences of bean yellow mosaic virus and two genomic sequences of clover yellow vein virus. The authors found recombinant events using seven RDP methods with P -values of 1.49×10^{-30} (HC-Pro) and 1.968×10^{-160} (CI-6k2). Recently, Peng et al. (2021) analyzed ten complete sequences of zucchini tigre mosaic potyvirus (ZTMV) and three papaya ringspot potyvirus (PRSV) isolates for recombination events using RDP4. Strong signals of independent recombination with seven RDP methods were found with calculated P -values of 1.3×10^{-11} - 3.1×10^{-57} (intraspecific recombination value of p) and 2.6×10^{-11} - 3.9×10^{-151} (interspecific recombination value of p).

In this study, RDP analysis with YMMV_DA_BRA_2019 calculated the following P -values for recombination events, RDP (4.672×10^{-17}), BootScan (1.87×10^{-16}), MaxChi (1.60×10^{-09}), Chimaera (6.61×10^{-08}), and 3seq (4.71×10^{-10}), and therefore, we did not consider these P -values significant. Moreover, the recombination region was only of 411 bp size with the beginning breaking point at position 8384 and ending breakpoint at 8795 in the YMMV_DA genomic sequence (Table 6).

388 **Table 6** Recombinant events detected in YMMV sequences determined by RDP software version 4.

Event No.	Recombinant	Minor Parental	Major Parental	Detection method ^a						
				R	G	B	M	C	S	3
1	^YMMV_DA_BRA_2019	Unknown (KJ125479_CHI_ala_2012)	YMMV_Dsp_BRA_2014	4.67E-17	NS	1.87E-16	1.60E-09	6.61E-08	NS	4.71E-10
1		Unknown (KJ125475_CHI_sp_2011)								
1		Unknown (KJ125478_CHI_ala_2011)								
2	^KJ125477_CHI_ala_2011	KC473517_CHI_pol_2010	KJ125474_CHI_sp_2011	NS	NS	NS	9.91E-10	4.07E-02	8.81E-04	3.18E-14
3	KC407674_CHI_opp_2011	KJ125473_CHI_ala_2010	KC473517_CHI_pol_2010	2.16E-09	NS	5.95E-11	1.83E-15	2.76E-08	5.50E-19	3.18E-14
4	^KX156847_SKO_pol_2014	KJ125478_CHI_ala_2011	YMMV_DA_BRA_2019	NS	NS	NS	1.39E-02	7.56E-03	NS	1.69E-06
4		KJ125475_CHI_sp_2011								
4		KJ125479_CHI_ala_2012								
5	^NC_019412_BRA_tri_2009	YMMV_DA_BRA_2019	Unknown (KX156847_SKO_pol_2014)	5.51E-20	4.29E-19	1.10E-18	2.62E-08	5.45E-12	5.25E-19	8.10E-09
5		YMMV_Dsp_BRA_2014								
6	^YMMV_Dsp_BRA_2014	KC407674_CHI_opp_2011	YMMV_DA_BRA_2019	NS	NS	NS	3.80E-07	3.44E-02	5.03E-50	NS
7	^KC473517_CHI_pol_2010	KJ125479_CHI_ala_2012	KJ125473_CHI_ala_2010	NS	NS	NS	NS	NS	1.33E-19	6.65E-04
7		KJ125475_CHI_sp_2011	KJ125476_CHI_ala_2011							
7		KJ125478_CHI_ala_2011								
8	^YMMV_Dsp_BRA_2014	KJ125474_CHI_sp_2011	KJ125478_CHI_ala_2011	NS	NS	NS	NS	3.74E-02	0.0128953	NS
8	YMMV_DA_BRA_2019		KJ125475_CHI_sp_2011							
9	^YMMV_DA_BRA_2019	Unknown (KJ125479_CHI_ala_2012)	YMMV_Dsp_BRA_2014	3.19E-03	1.38E-03	NS	NS	NS	NS	NS
9		Unknown (KJ125475_CHI_sp_2011)								
9		Unknown (KJ125478_CHI_ala_2011)								

389 ^a Method abbreviation = R: RDP, G: GENECONV, B: Bootscan, M: Maxchi, C: Chimaera, S: SiScan, and 3: 3seq.

Table Key:

~ = It is possible that this apparent recombination signal could have been caused by an evolutionary process other than recombination.

* = The actual breakpoint position is undetermined (it was most likely overprinted by a subsequent recombination event).

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

Minor Parent = Parent contributing the smaller fraction of sequence.

Major Parent = Parent contributing the larger fraction of sequence.

Unknown = Only one parent and a recombinant need be in the alignment for a recombination event to be detectable.

The sequence listed as unknown was used to infer the existence of a missing parental sequence.

NS = No significant P-value was recorded for this recombination event using this method.

390

391 **4. Conclusions**

392 This study reveals the presence of YMMV infections only in *D. alata* plants from fields
 393 at the northeast (Pernambuco and Paraíba) of Brazil. No infection with YMV was found in *D.*
 394 *alata*. HTS and ELISA data did not reveal mixed infections of YMMV and YMV. Bayesian
 395 inference of phylogeny using 67 nt sequences of the CP region restructured the phylogenetic
 396 groups of YMMV into ten phylogroups (I-X), in which the YMMV sequences of this study
 397 were located within group IX. This phylogenetic analysis suggests that there is no correlation
 398 between YMMV isolates and geographic location. Last, recombination analysis based on 14
 399 YMMV completed genomic sequences using RDP4 did not reveal any recombination event
 400 with significant *P*-values. This work has shed light on the molecular characterization and
 401 phylogenetic relationships among YMMV isolates around the world.

Acknowledgments

The authors acknowledge the financial support given by CNPq (Project 408709/2018-9), CAPES-PrInt-UFRPE, and the Rural Federal University of Pernambuco (Edital 015/2018, UFRPE-PRPPG).

Thanks to the yam farmers who allowed access to their fields to collect samples of study material. We are also thankful for providing samples (yam tubers) and assistance in sample collection in non-cultivated areas: MSc. Gabriel Pavan Sabino (Universidade Estadual Paulista (UNESP), Rio Claro, São Paulo), Dr. Diogo Amorim de Araújo (Universidade Federal de Pernambuco-UFPE), Pernambuco, Dr. Jones Clebson R. Mendez (Universidade Federal Rural de Pernambuco-UFRPE).

Thanks to researchers of the Agronomic Institute of Pernambuco (IPA): Dr. José Nildo Tabosa, and MSc. Almir Dias Alves da Silva for providing us with access to yam's production fields.

Funding

This work was supported by funding from CNPq (Project 408709/2018-9), CAPES-PrInt-UFRPE, and the Rural Federal University of Pernambuco (Edital 015/2018, UFRPE-PRPPG).

Data availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Consent for publication All contributing authors have read and approved of the final version of the manuscript.

430 **References**

- 431 Adams MJ, Antoniw JF, Beaudoin F (2005) Overview and analysis of the polyprotein
432 cleavage sites in the family Potyviridae. *Mol Plant Pathol* 6:471–487.
433 <https://doi.org/10.1111/j.1364-3703.2005.00296.x>
- 434 Adeniji MO, Shoyinka SA, Ikotun T, et al (2012) Yield loss in guinea yam (*Dioscorea*
435 *rotundata* poir.) due to infection by yam mosaic virus (YMV) genus potyvirus. *Ife J*
436 *Sci* 14:237–244
- 437 Aleman ME, Marcos JF, Brugidou C, et al (1996) The complete nucleotide sequence of
438 yam mosaic virus (Ivory Coast isolate) genomic RNA. *Arch Virol* 141:1259–1278.
439 <https://doi.org/10.1007/BF01718829>
- 440 Atreya PL, Lopez-Moya JJ, Chu M, et al (1995) Mutational analysis of the coat protein N-
441 terminal amino acids involved in potyvirus transmission by aphids. *J Gen Virol*
442 76:265–270. <https://doi.org/10.1099/0022-1317-76-2-265>
- 443 Bankevich A, Nurk S, Antipov D, et al (2012) SPAdes: A new genome assembly algorithm
444 and its applications to single-cell sequencing. *J Comput Biol* 19:455–477.
445 <https://doi.org/10.1089/cmb.2012.0021>
- 446 Blanc S, Ammar ED, Garcia-Lampasona S, et al (1998) Mutations in the potyvirus helper
447 component protein: Effects on interactions with virions and aphid stylets. *J Gen Virol*
448 79:3119–3122. <https://doi.org/10.1099/0022-1317-79-12-3119>
- 449 Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina
450 sequence data. *Bioinformatics* 30:2114–2120.
451 <https://doi.org/10.1093/bioinformatics/btu170>
- 452 Bömer M, Rathnayake AI, Visendi P, et al (2018) Complete genome sequence of a new
453 member of the genus Badnavirus, *Dioscorea bacilliform* RT virus 3, reveals the first
454 evidence of recombination in yam badnaviruses. *Arch Virol* 163:533–538.
455 <https://doi.org/10.1007/s00705-017-3605-9>
- 456 Bömer M, Turaki AA, Silva G, et al (2016) A sequence-independent strategy for
457 amplification and characterisation of episomal badnavirus sequences reveals three
458 previously uncharacterised yam badnaviruses. *Viruses* 8:14–16.
459 <https://doi.org/10.3390/v8070188>
- 460 Bousalem M, Dallot S, Fuji S, Natsuaki KT (2003) Origin, world-wide dispersion, bio-
461 geographical diversification, radiation and recombination: An evolutionary history of
462 Yam mild mosaic virus (YMMV). *Infect Genet Evol* 3:189–206.
463 [https://doi.org/10.1016/S1567-1348\(03\)00085-6](https://doi.org/10.1016/S1567-1348(03)00085-6)
- 464 Bousalem M, Dallot S, Guyader S (2000) The use of phylogenetic data to develop
465 molecular tools for the detection and genotyping of Yam mosaic virus. Potential
466 application in molecular epidemiology. *J Virol Methods* 90:25–36.
467 [https://doi.org/10.1016/S0166-0934\(00\)00198-1](https://doi.org/10.1016/S0166-0934(00)00198-1)

- 468 Braidwood L, Müller SY, Baulcombe D (2019) Extensive recombination challenges the
469 utility of Sugarcane mosaic virus phylogeny and strain typing. *Sci Rep* 9:1–9.
470 <https://doi.org/10.1038/s41598-019-56227-y>
- 471 Chai M, Wu X, Liu J, et al (2020) P3N-PIPO Interacts with P3 via the Shared N-Terminal
472 Domain To Recruit Viral Replication Vesicles for Cell-to-Cell Movement. *J Virol*
473 94:. <https://doi.org/10.1128/jvi.01898-19>
- 474 Dallot S, Guzmán M, Bousalem M (2001) Occurrence of Potyviruses on Yam (*Dioscorea*
475 spp.) in Colombia and First Molecular Characterization of Yam mild mosaic virus .
476 *Plant Dis* 85:803–803. <https://doi.org/10.1094/pdis.2001.85.7.803d>
- 477 Darriba D, Taboada GL, Doallo R, Posada D (2012) JModelTest 2: More models, new
478 heuristics and parallel computing. *Nat Methods* 9:772.
479 <https://doi.org/10.1038/nmeth.2109>
- 480 Desbiez C, Wipf-Scheibel C, Millot P, et al (2017) New species in the papaya ringspot
481 virus cluster: Insights into the evolution of the PRSV lineage. *Virus Res* 241:88–94.
482 <https://doi.org/10.1016/j.virusres.2017.06.022>
- 483 Eni AO, Hughes JDA, Asiedu R, Rey MEC (2010) Survey of the incidence and
484 distribution of viruses infecting yam (*Dioscorea* spp.) in Ghana and Togo. *Ann Appl*
485 *Biol* 156:243–251. <https://doi.org/10.1111/j.1744-7348.2009.00383.x>
- 486 Eni AO, Lava Kumar P, Asiedu R, et al (2013) Characterization of cucumber mosaic virus
487 isolated from yam (*Dioscorea* spp.) in West Africa. *African J Biotechnol* 12:3472–
488 3480. <https://doi.org/10.5897/AJB2013.12303>
- 489 Filloux D, Bonheur L, Umber M, et al (2015) Metagenomic discovery, worldwide
490 distribution and genetic diversity of novel macluraviruses infecting yams (*Dioscorea*
491 spp.). 15èmes Rencontres Virol Végétale
- 492 Fuji S, Mitobe I, Nakamae H, Natsuaki KT (1999) Nucleotide sequence of coat protein
493 gene of yam mild mosaic virus, isolated in Papua New Guinea. *Arch Virol* 144:1415–
494 1419. <https://doi.org/10.1007/s007050050597>
- 495 Fuji S, Nakamae H (1999) Complete nucleotide sequence of the geonomic RNA of a mild
496 strain of Japanese yam mosaic potyvirus. *Arch Virol* 145:635–640.
497 <https://doi.org/10.1007/s007050050052>
- 498 Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-scanning: A Monte Carlo procedure for
499 assessing signals in rebombinant sequences. *Bioinformatics* 16:573–582.
500 <https://doi.org/10.1093/bioinformatics/16.7.573>
- 501 Hasegawa M, Kishino H, Yano T aki (1985) Dating of the human-ape splitting by a
502 molecular clock of mitochondrial DNA. *J Mol Evol* 22:160–174.
503 <https://doi.org/10.1007/BF02101694>
- 504 Hayashi EAI, Blawid R, de Melo FL, et al (2017) Complete genome sequence of a putative
505 new secovirus infecting yam (*Dioscorea*) plants. *Arch Virol* 162:317–319.
506 <https://doi.org/10.1007/s00705-016-3104-4>

- 507 Huet H, Gal-On A, Meir E, et al (1994) Mutations in the helper component protease gene
508 of zucchini yellow mosaic virus affect its ability to mediate aphid transmissibility. J
509 Gen Virol 75:1407–1414. <https://doi.org/10.1099/0022-1317-75-6-1407>
- 510 Karavina C, Ibaba JD, Gubba A (2020) High-throughput sequencing of virus-infected
511 Cucurbita pepo samples revealed the presence of Zucchini shoestring virus in
512 Zimbabwe. BMC Res Notes 13:1–6. <https://doi.org/10.1186/s13104-020-4927-3>
- 513 Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7:
514 Improvements in performance and usability. Mol Biol Evol 30:772–780.
515 <https://doi.org/10.1093/molbev/mst010>
- 516 Kenyon L, Shoyinka SA, Hughes J d'A., Odu BO (2001) An overview of viruses infecting
517 Dioscorea yams in sub-Saharan Africa. Plant Virol Sub-Saharan Africa 432–439
- 518 Kondo T, Fujita T (2012) Complete nucleotide sequence and construction of an infectious
519 clone of Chinese yam necrotic mosaic virus suggest that macluraviruses have the
520 smallest genome among members of the family Potyviridae. Arch Virol 157:2299–
521 2307. <https://doi.org/10.1007/s00705-012-1429-1>
- 522 Lam HM, Ratmann O, Boni MF (2018) Improved Algorithmic Complexity for the 3SEQ
523 Recombination Detection Algorithm. Mol Biol Evol 35:247–251.
524 <https://doi.org/10.1093/molbev/msx263>
- 525 Lan P, Meng Y, Shen P, et al (2018) Complete genome sequence of yam chlorotic necrosis
526 virus, a novel macluravirus infecting yam. Arch Virol 163:2275–2278.
527 <https://doi.org/10.1007/s00705-018-3851-5>
- 528 Lan P, Tan G, Li F Identification and characterization of a novel potyvirus infecting Paris
529 yunnanensis. 1–8
- 530 Mambole IA, Bonheur L, Dumas LS, et al (2014) Molecular characterization of yam virus
531 X, a new potexvirus infecting yams (Dioscorea spp) and evidence for the existence
532 of at least three distinct potexviruses infecting yams. Arch Virol 159:3421–3426.
533 <https://doi.org/10.1007/s00705-014-2211-3>
- 534 Marais A, Umber M, Filloux D, et al (2020) Yam asymptomatic virus 1, a novel virus
535 infecting yams (Dioscorea spp.) with significant prevalence in a germplasm
536 collection. Arch Virol 165:2653–2657. <https://doi.org/10.1007/s00705-020-04787-0>
- 537 Martin DP, Murrell B, Golden M, et al (2015) RDP4: Detection and analysis of
538 recombination patterns in virus genomes. Virus Evol 1:1–5.
539 <https://doi.org/10.1093/ve/vev003>
- 540 Martin DP, Posada D, Crandall KA, Williamson C (2005) A modified bootscan algorithm
541 for automated identification of recombinant sequences and recombination
542 breakpoints. AIDS Res Hum Retroviruses 21:98–102.
543 <https://doi.org/10.1089/aid.2005.21.98>
- 544 Menzel W, Thottappilly G, Winter S (2014) Characterization of an isometric virus isolated
545 from yam (Dioscorea rotundata) in Nigeria suggests that it belongs to a new species

- 546 in the genus *Aureusvirus*. *Arch Virol* 159:603–606. [https://doi.org/10.1007/s00705-](https://doi.org/10.1007/s00705-013-1845-x)
547 013-1845-x
- 548 Muhire BM, Varsani A, Martin DP (2014) SDT: A virus classification tool based on
549 pairwise sequence alignment and identity calculation. *PLoS One* 9:.
550 <https://doi.org/10.1371/journal.pone.0108277>
- 551 Mumford RA, Seal SE (1997) Rapid single-tube immunocapture RT-PCR for the detection
552 of two yam potyviruses. *J Virol Methods* 69:73–79. [https://doi.org/10.1016/S0166-](https://doi.org/10.1016/S0166-0934(97)00141-9)
553 0934(97)00141-9
- 554 Nkere CK, Otoo E, Atiri GI, et al (2020) Assessment of Yam mild mosaic virus coat
555 protein gene sequence diversity reveals the prevalence of cosmopolitan and African
556 group of isolates in Ghana and Nigeria. *Curr Plant Biol* 23:100156.
557 <https://doi.org/10.1016/j.cpb.2020.100156>
- 558 Odu BO, Hughes JDA, Shoyinka SA, Dongo LN (1999) Isolation, characterisation and
559 identification of a potyvirus from *Dioscorea alata* L. (water yam) in Nigeria. *Ann*
560 *Appl Biol* 134:65–71. <https://doi.org/10.1111/j.1744-7348.1999.tb05236.x>
- 561 Ohshima K, Tomitaka Y, Wood JT, et al (2007) Patterns of recombination in turnip mosaic
562 virus genomic sequences indicate hotspots of recombination. *J Gen Virol* 88:298–
563 315. <https://doi.org/10.1099/vir.0.82335-0>
- 564 Ohshima K, Yamaguchi Y, Hirota R, et al (2002) Molecular evolution of Turnip mosaic
565 virus: Evidence of host adaptation, genetic recombination and geographical spread. *J*
566 *Gen Virol* 83:1511–1521. <https://doi.org/10.1099/0022-1317-83-6-1511>
- 567 Olspert A, Chung BY, Atkins JF, et al (2015) Transcriptional slippage in the positive-
568 sense RNA virus family Potyviridae . *EMBO Rep* 16:995–1004.
569 <https://doi.org/10.15252/embr.201540509>
- 570 Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by
571 frequent recombination. *Virology* 265:218–225.
572 <https://doi.org/10.1006/viro.1999.0056>
- 573 Peng B, Liu L, Wu H, et al (2021) Interspecific Recombination Between Zucchini Tigre
574 Mosaic Virus and Papaya Ringspot Virus Infecting Cucurbits in China. *Front*
575 *Microbiol* 12:. <https://doi.org/10.3389/fmicb.2021.773992>
- 576 Peng YH, Kadoury D, Gal-On A, et al (1998) Mutations in the HC-Pro gene of zucchini
577 yellow mosaic potyvirus: Effects on aphid transmission and binding to purified
578 virions. *J Gen Virol* 79:897–904. <https://doi.org/10.1099/0022-1317-79-4-897>
- 579 Pio-Ribeiro G, Andrade GP, Filloux D, et al (2006) Plantas de inhame em Pernambuco e
580 Paraíba apresentam infecções simples e mista por potyvírus e badnavírus. *Fitopatol*
581 *Bras v.* 31:309
- 582 Pirone TP, Blanc S (1996) Helper-dependent vector transmission of plant viruses. *Annu*
583 *Rev Phytopathol* 34:227–247. <https://doi.org/10.1146/annurev.phyto.34.1.227>

- 584 Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from
585 DNA sequences: Computer simulations. *Proc Natl Acad Sci U S A* 98:13757–13762.
586 <https://doi.org/10.1073/pnas.241370698>
- 587 Rabelo-Filho F de AC, Nicolini C, Resende R de O, et al (2013) The complete genome
588 sequence of a Brazilian isolate of yam mild mosaic virus. *Arch Virol* 158:515–518.
589 <https://doi.org/10.1007/s00705-012-1509-2>
- 590 Rambaut A, Drummond AJ, Xie D, et al (2018) Posterior summarization in Bayesian
591 phylogenetics using Tracer 1.7. *Syst Biol* 67:901–904.
592 <https://doi.org/10.1093/sysbio/syy032>
- 593 Revers F, García JA (2015) Molecular biology of potyviruses. *Adv. Virus Res.* 92:101–
594 199
- 595 Ronquist F, Huelsenbeck J, Teslenko M (2011) MrBayes Version 3.2 Manual: Tutorials
596 and Model Summaries. *Man MrBayes* 1–103
- 597 Seal S, Muller E (2007) Molecular analysis of a full-length sequence of a new yam
598 badnavirus from *Dioscorea sansibarensis*. *Arch Virol* 152:819–825.
599 <https://doi.org/10.1007/s00705-006-0888-7>
- 600 Shibolet Y, Haronsky E, Leibman D, et al (2007) The Conserved FRNK Box in HC-
601 Pro, a Plant Viral Suppressor of Gene Silencing, Is Required for Small RNA Binding
602 and Mediates Symptom Development. *J Virol* 81:13135–13148.
603 <https://doi.org/10.1128/jvi.01031-07>
- 604 Silva G, Bömer M, Rathnayake AI, et al (2019) Molecular Characterization of a New Virus
605 Species Identified in Yam (*Dioscorea* spp.) by High-Throughput Sequencing. *Plants*
606 8:167. <https://doi.org/10.3390/plants8060167>
- 607 Smith JM (1992) Analyzing the mosaic structure of genes. *J Mol Evol* 34:126–129.
608 <https://doi.org/10.1007/BF00182389>
- 609 Su X, Fu S, Qian Y, et al (2016) Discovery and small RNA profile of Pecan mosaic-
610 associated virus, a novel potyvirus of pecan trees. *Sci Rep* 6:1–7.
611 <https://doi.org/10.1038/srep26741>
- 612 Sukal A, Kidanemariam D, Dale J, et al (2017) Characterization of badnaviruses infecting
613 *Dioscorea* spp. in the Pacific reveals two putative novel species and the first report of
614 *dioscorea bacilliform RT virus* 2. *Virus Res* 238:29–34.
615 <https://doi.org/10.1016/j.virusres.2017.05.027>
- 616 Torrance L, Andreev IA, Gabrenaite-Verhovskaya R, et al (2006) An unusual structure at
617 one end of potato potyvirus particles. *J Mol Biol* 357:1–8.
618 <https://doi.org/10.1016/j.jmb.2005.12.021>
- 619 Umber M, Gomez RM, Gélabale S, et al (2017) The genome sequence of *Dioscorea*
620 bacilliform TR virus, a member of the genus Badnavirus infecting *Dioscorea* spp.,
621 sheds light on the possible function of endogenous *Dioscorea bacilliform* viruses.
622 *Arch Virol* 162:517–521. <https://doi.org/10.1007/s00705-016-3113-3>

- 623 Uzeyir M, Gargani D, Drucker M, et al (2007) A protein key to plant virus transmission at
624 the tip of the insect vector stylet. *Proc Natl Acad Sci U S A* 104:17959–17964.
625 <https://doi.org/10.1073/pnas.0706608104>
- 626 Wang M, Li F, Zhou G, et al (2015) Molecular Detection and Characterization of Chinese
627 Yam Mild Mosaic Virus Isolates. *J Phytopathol* 163:1036–1040.
628 <https://doi.org/10.1111/jph.12337>
- 629 Worrall EA, Hayward AC, Fletcher SJ, Mitter N (2019) Molecular characterization and
630 analysis of conserved potyviral motifs in bean common mosaic virus (BCMV) for
631 RNAi-mediated protection. *Arch Virol* 164:181–194.
632 <https://doi.org/10.1007/s00705-018-4065-6>
- 633 Wylie SJ, Adams M, Chalam C, et al (2017) ICTV virus taxonomy profile: Potyviridae. *J*
634 *Gen Virol* 98:352–354. <https://doi.org/10.1099/jgv.0.000740>
- 635 Wylie SJ, Kueh J, Welsh B, et al (2002) A non-aphid-transmissible isolate of bean yellow
636 mosaic potyvirus has an altered NAG motif in its coat protein. *Arch Virol* 147:1813–
637 1820. <https://doi.org/10.1007/s00705-002-0846-y>
- 638 Yu XQ, Lan YF, Wang HY, et al (2007) The complete genomic sequence of Tobacco vein
639 banding mosaic virus and its similarities with other potyviruses. *Virus Genes* 35:801–
640 806. <https://doi.org/10.1007/s11262-007-0135-7>
- 641 Zhang P, Peng J, Guo H, et al (2016) Complete genome sequence of yam chlorotic necrotic
642 mosaic virus from *Dioscorea parviflora*. *Arch Virol* 161:1715–1717.
643 <https://doi.org/10.1007/s00705-016-2818-7>
- 644 Zou C, Meng J, Li Z, et al (2011) First Report of Yam mild mosaic virus in Yam in
645 Guangxi Province, China. *Plant Dis* 95:1320–1320. <https://doi.org/10.1094/pdis-04-11-0350>
- 647 Zou CW, Meng JR, Yao ZT, et al (2020) Genetic diversity and genome recombination in
648 Yam mild mosaic virus isolates. *Phytopathol Res* 2:. <https://doi.org/10.1186/s42483-020-00051-0>
649
- 650
- 651
- 652

CONCLUSÕES GERAIS



CONCLUSÕES GERAIS

1- Seis espécies de plantas de *Dioscorea* spp. foram coletadas e agrupados em *pools* amostrais, DR1 (*Dioscorea cayennensis-rotundata*, Alagoas), DR2 (*D. cayennensis-rotundata*, Pernambuco, Paraíba e Alagoas), DB (*D. bulbifera*, Pernambuco e São Paulo) e ID (*D. trifida*, *D. subhastata* e *D. altissima*, Pernambuco). Os dados de HTS permitiram a montagem das seguintes sequências genômicas virais de: (i) yam mosaic virus (*Potyvirus*), dioscorea bacilliform AL virus (*Badnavirus*) e dioscorea mosaic-associated virus (RNA1 e RNA2, *Sadwavirus*) dos *pools* DR1 e DR2, e adicionalmente dioscorea baciliforme RT virus (*Badnavirus*) do *pool* DR2; (ii) yam mild mosaic virus (*Potyvirus*), dioscorea bacilliform AL virus (*Badnavirus*) e dioscorea virus A (Família *Betaflexiviridae*) do *pool* DA; (iii) yam mosaic virus e yam mild mosaic virus do *pool* Dsp. Nenhuma sequência viral foi encontrada no *pool* ID.

2- Os resultados de ELISA mostraram que o YMV está presente apenas em amostras de *D. cayennensis-rotundata* de todos os estados de amostragem (Alagoas, Bahia, Maranhão, Paraíba e Pernambuco), enquanto o YMMV foi detectado apenas em *D. alata* (Paraíba e Pernambuco). Estes resultados estão de acordo com os dados obtidos de HTS, onde dos *pools* DR1 e DR2 (*D. cayennensis-rotundata*) foi possível montar apenas sequências genômicas de YMV e do *pool* DA (*D. alata*) apenas a sequência genômica de YMMV. Nenhuma das plantas de *D. bulbifera* apresentou resultado positivo nos testes de ELISA para infecção com estes dois vírus. Também não foi possível achar sequência de YMMV e YMV nos dados de HTS do *pool* DB.

3- Três sequências genômicas de YMV foram montadas dos *pools* DR1 (YMV_DR1), DR2 (YMV_DR2) e Dsp (YMV_Dsp). Uma sequência completa genômica de YMV (YMV_I4) foi validada por sequenciamento Sanger.

4- A análise filogenética bayesiana usando sequências da região codificadora do gene CP revelou que os isolados YMV_DR1, YMV_DR2, YMV_Dsp e YMV_I4 formaram um novo filogrupo (Grupo X), diferente dos relatados até o presente momento. A análise filogenética temporal usando sequências nucleotídicas de CP de YMV estimou: (i) o melhor modelo de relógio o relaxado lognormal não correlacionado (UCLN), (ii) uma taxa evolutiva média $1,76 \times 10^{-3}$ substituições por sítio por ano, similar às encontradas com vírus de RNA, (iii) o tempo médio estimado do ancestral comum mais recente (TMRCA) em 168,68 anos atrás, e (iv) a

origem geográfica mais provável foi no continente africano, mais especificamente no país da Nigéria.

5- A sequência genômica de YMMV_DA (*pool* DA) consistiu em 9.512 nt e a sequência de YMMV_Dsp (*pool* Dsp) em 9.501 nt, excluindo a cauda poly(A). Os isolados YMMV_DA e YMMV_Dsp são filogeneticamente distantes do isolado brasileiro de YMMV de *D. trifida*, e, portanto, compartilham uma identidade de sequência nucleotídica relativamente baixa (~87%). A análise Bayesiana realizada com 67 sequências nt de CP resultou na reestruturação dos grupos filogenéticos de YMMV em dez filogrupos (I-X). As sequências de YMMV_DA e YMMV_Dsp foram agrupadas no *cluster* IX (hospedeira *D. alata* e isolados de Togo, Nigéria e Brasil), e portanto, concluiu-se que não há correlação com a distribuição geográfica dos isolados. Finalmente, a análise de recombinação com base em 14 sequências genômicas completas de YMMV usando RDP4 não revelou nenhum evento significativo de recombinação.