



UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS NATURAIS E
BIOTECNOLOGIA

APLICAÇÃO E VALIDAÇÃO DE METODOLOGIA PARA IDENTIFICAÇÃO E
CARACTERIZAÇÃO MOLECULAR DE VÍRUS ASSOCIADO À ATROFIA DA
COROA DO COQUEIRO

Henriqueta Monalisa Farias

Cuité

2022

Henriqueta Monalisa Farias

**APLICAÇÃO E VALIDAÇÃO DE METODOLOGIA PARA IDENTIFICAÇÃO E
CARACTERIZAÇÃO MOLECULAR DE VÍRUS ASSOCIADO À ATROFIA DA
COROA DO COQUEIRO**

Dissertação realizada sob orientação da Dra. Simone da Graça Ribeiro, apresentada ao Programa de Pós-Graduação em Ciências Naturais e Biotecnologia do Centro de Educação e Saúde da Universidade Federal de Campina Grande para obtenção do título de Mestre.

Orientador: Dra. Simone da Graça Ribeiro

Coorientador: Prof. Dra. Magnólia de Araújo Campos

Cuité

2022

F224a

Farias, Henriqueta Monalisa.

Aplicação e validação de metodologia para identificação e caracterização molecular de vírus associado à Atrofia da Coroa do Coqueiro. / Henriqueta Monalisa Farias. - Cuité, 2022.

77 f. : il. color.

Dissertação (Mestrado em Ciências Naturais e Biotecnologia) - Universidade Federal de Campina Grande, Centro de Educação e Saúde, 2022.

"Orientação: Profa. Dra. Simone da Graça Ribeiro; Profa. Dra. Magnólia de Araújo Campos".

Referências.

1. Coco. 2. Coqueiro. 3. Atrofia da coroa do coqueiro. 4. Atrofia da coroa do coqueiro - identificação - vírus. 5. Coqueiro - doença - vírus. 6. *Nanoviridae*. 7. *Cocos nucifera* L. I. Ribeiro, Simone da Graça . II. Campos, Magnólia de Araújo . III. Título.

CDU 634.616(043)

APLICAÇÃO E VALIDAÇÃO DE METODOLOGIA PARA IDENTIFICAÇÃO E CARACTERIZAÇÃO MOLECULAR DE VÍRUS ASSOCIADO À ATROFIA DA COROA DO COQUEIRO

HENRIQUETA MONALISA FARIAS

Dissertação aprovada em 11/03/2022 pela Banca Examinadora constituída dos seguintes membros:

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE EDUCAÇÃO E SAÚDE
MESTRADO EM CIÊNCIAS NATURAIS E BIOTECNOLOGIA



Profa. Dra. Simone da Graça Ribeiro
Orientadora



Profa. Dra. Magnólia de Araújo Campos Pfenning
Coorientadora



Prof. Dr. Rafael Trindade Maia
Examinador Interno



Henriqueta Monalisa Farias
Candidata



Profa. Dra. Alessandra de Jesus Boari
Examinadora Externa



Profa. Dra. Magnólia de Araújo Campos Pfenning
Coordenadora do PPG-CNBiotec

AGRADECIMENTOS

Dentre tantas coisas, agradecer a Deus é a principal delas. Foi ele quem me deu forças, discernimento, conforto, me acalmou em momentos de ansiedade ou desespero. Eu te amo senhor, obrigado pelo dom da vida, pela minha saúde, por estar aqui comigo em todos os momentos, e essencialmente, por sempre manter meu coração vivo em esperança.

A minha família, meu pai, minha mãe, meus irmãos, meu padrasto e minha tia Santa. Vocês são a base mais importante da minha vida.

Em meras palavras não conseguiria agradecer ao apoio da minha orientadora, Simone Ribeiro, ela foi sem dúvidas uma das minhas principais motivadoras. Obrigada pela sua compreensão em tantos momentos difíceis, pela paciência e por tanto conhecimento adquirido nesses dois anos de convivência. Você tem o dom de repassar conhecimento, de agregar outras pessoas e contribuir com o mundo.

Ao Fernando Melo e Alessandra Boari por toda contribuição com a pesquisa e disposição para auxílio, sem eles nada disso seria possível.

A dona Neide, anjo que me acolheu tão bem em Brasília, com ela pessoas muito especiais para mim: Daiane, Layane e Pablo. Não tenho como agradecer por tanto apoio e acolhimento, amo vocês.

Aos amigos que a Embrapa me proporcionou, Thais, Andressa, Clara, Monique, Pedro, Gustavo, Bruna Lima, Bruna Medeiros, e especialmente, Andreza Vidal, agradeço o suporte, conversas construtivas e com carinho levarei para minha vida.

A Magnólia, amiga e coorientadora, uma mulher de fé. Obrigada por tantas palavras de conforto e paz, por sempre lembrar que Deus sabe de todas as coisas.

Ao PPG CNBiotec, programa que me proporcionou chegar até aqui. Aos amigos que fiz, Yam, Ruth, Rodrigo, Samara, Amanda, Maria Tereza, Alana e Mikael. Em especial Yam, esse caminhou junto comigo desde o início e se fez presente em diversos momentos, gratidão.

Por fim, e muito importante, o suporte de pessoas especiais na minha vida Leonardo, Luana, Esli, Helany, Joana e Rafael, meu muito obrigado.

O desenvolvimento deste trabalho exigiu de mim muita coragem e determinação para percorrer caminhos. Durante o período de dois anos precisei modificar substancialmente a

maioria dos segmentos da minha vida para fazer com que eles se realinhassem, a fim de atender exclusivamente à satisfação profissional de concluir o meu mestrado. Para isso se concretizar, enfrentei incalculáveis desafios, dentre eles, o mais árduo de todos: o de conseguir superar as adversidades potencializadas pela minha própria mente.

Deu(s) tudo certo.

“If you want to change the world in some big way, that’s where you should start – biological molecules.”

Bill Gates

RESUMO

No Brasil, o coqueiro é uma cultura de grande importância econômica. O país é o principal produtor mundial de água de coco e quinto maior produtor mundial de frutos do coqueiro. Assim como acontece em outras culturas, o coco também é suscetível a pragas e problemas fitossanitários que podem prejudicar seu cultivo. A atrofia da coroa do coqueiro (ACC), é uma doença que teve seu primeiro caso registrado em 2012 no Brasil. O nome da doença é devido a sintomatologia causada nas plantas, podendo ocasionar aborto dos frutos, necrose, assim como o encurtamento foliar. Somado a isso, a grande problemática em torno da atrofia da coroa do coqueiro está na falta de informações relacionadas a sua etiologia bem como possíveis vetores transmissores. Por meio de um sequenciamento de nova geração de amostras de coqueiros apresentando sintomas da doença, foi identificado um DNA-Rep semelhante a DNA-Reps associados a vírus membros da família *Nanoviridae*. Nesse contexto, o presente trabalho teve como objetivo principal, descrever e validar uma metodologia de identificação de um possível vírus multipartido associado a atrofia da coroa do coqueiro (coconut crown atrophy virus - CCAV) a partir de *primers* permitindo acesso a diversidade molecular viral fazendo uso de sequenciamento nanopore. Para isso, foram utilizadas amostras de coqueiro de cinco estados do Brasil, apresentando ou não sintomas da doença. Com base na sequência da região comum maior (CR-M) de componentes do CCAV previamente clonados e sequenciados, foram desenhados pares de primers para amplificar todos os componentes genômicos do CCAV em uma reação de PCR inverso multiplex. Posteriormente, foi realizado o sequenciamento dos *amplicons* obtidos no sequenciador MinION Mk1B (Oxford Nanopore Technologies). Os primers CR1-Mix utilizados em PCR multiplex se mostraram eficientes no acesso a diversidade genética evidenciando sete componentes genômicos (DNA-1, 2, 3, 4, 5, 6 e 7) dos isolados de coconut crown atrophy virus (CCAV), especialmente dos isolados coletados nos estados do Pará. A PCR multiplex associada ao sequenciamento nanopore mostrou-se como uma ferramenta rápida, eficaz e assertiva para caracterização molecular do CCAV frente a outras técnicas de clonagem e sequenciamento. Por fim, os estudos filogenéticos sugerem que o CCAV é um membro de um novo táxon podendo representar um novo gênero da família *Nanoviridae* ou um membro de uma nova família de vírus de genoma de ssDNA de plantas.

Palavras-Chave: *Nanoviridae*, *Cocos nucifera* L., nanopore sequencing, coconut crown atrophy

ABSTRACT

Coconut is a crop of great economic importance in Brazil. The country is the world's leading producer of coconut water and the world's fifth largest producer of coconut fruit. As with other crops, coconut is also susceptible to pests and phytosanitary problems that can affect the production. Coconut crown atrophy (CCA) was first reported in 2012 in Brazil. The disease name is attributed to the symptoms induced in the plants, which can cause fruit loss, necrosis, and leaf atrophy. CCA etiology and its possible vectors are hitherto unknown. High throughput sequencing studies of coconut palms showing disease symptoms identified a DNA-Rep sequence similar to DNA-Rep associated with viruses belonging to the *Nanoviridae* family. In the present study, we describe and validate methods for the identification of coconut crown atrophy virus – CCAV, a possible multipartite virus associated with CCA. We also standardize a nanopore sequencing method to access CCAV diversity. Palm samples from five Brazilian states were collected from symptomatic and asymptomatic coconut plants. The sequence of the largest common region (CR-M) of previously cloned and sequenced CCAV components was used to design primer pairs to amplify all genomic components of CCAV in a multiplex inverse PCR reaction. Then, the amplicons were sequenced in a MinION Mk1B sequencer (Oxford Nanopore Technologies). The CR1-Mix primers used in multiplex PCR proved to be efficient in accessing the genetic diversity of the coconut crown atrophy virus (CCAV) genomes of the studied isolates, especially from Pará state. Multiplex PCR associated with nanopore sequencing proved to be a fast, effective and assertive tool for molecular characterization of CCAV when compared to other cloning and sequencing techniques. Finally, phylogenetic studies suggest that CCAV is a member of a new taxon that could represent a new genus within the *Nanoviridae* family or a member of a new family of plant ssDNA genome viruses.

Palavras-Chave: *Nanoviridae*, *Cocos nucifera* L, nanopore sequencing, coconut crown atrophy

LISTA DE FIGURAS

Figura 1 - Coqueiro apresentando a doença atrofia da coroa do coqueiro. Fonte: Boari, dados não publicados.	22
Figura 2 - Estrutura do capsídeo dos nanovírus em forma icosaédrica contendo o ssDNA em seu interior. Fonte: (VIRALZONE, 2009; https://viralzone.expasy.org/125?outline=all_by_species).	24
Figura 3 - Componentes genômicos dos nanovírus DNA-R, S, C, M, N, U1, U2 e U4 com suas respectivas proteínas codificadas; CR-SL, região comum I (região formadora do Steam Loop, M-rep reconhecem a origem de replicação); CR-II segunda região de sequências conservadas; ORF, open Reading frame (início de região codificadora de proteínas ou RNAs funcionais para sequência de DNA); TATA Box, sequência promotora reconhecida pelos fatores de transcrição; Poli A, sinal de poliadenilação. Fonte: Imagem feita pela autora de acordo com informações do ICTV sobre o genoma do nanovírus faba bean necrotic yellows virus (FBNYV).	26
Figura 4 - Esquematização da tecnologia de sequenciamento nanopore da Oxford Nanopore Technologies. Fonte: Adaptado https://nanoporetech.com/	32
Figura 5 - Eletroforese em gel de agarose dos produtos da PCR com os primers BABUR e BABUF do RCA isolado das amostras de coco 4 e 12 do Pará. M- Marcador de peso molecular 1 Kb plus; C- Controle negativo.	45
Figura 6 - Eletroforese em gel de agarose evidenciando padrão de restrição da digestão com a enzima Eco RI dos clones recombinantes dos isolados 04 e 12 em pCR2.1_TOPO. M- Marcador de peso molecular 1 Kb plus.	46
Figura 7 - Alinhamento da sequência de nucleotídeos dos clones virais 4 e 12. Em vermelho são evidenciados os nucleotídeos com 100% de identidade e ao lado representação dos componentes Rep, em verde encontram-se as ORF's determinadas pelo programa ORFfinder.	47
Figura 8 - Sequências dos componentes genômicos DNA-1, 2, 3, 4, 6 e 7 utilizadas no desenho dos primers para PCR multiplex. Fonte: Autora.	49
Figura 9 - Eletroforese em gel de agarose de fragmentos resultantes da RCA-PCR com os primers CR1-MIX com DNAs isolados das amostras de coqueiro Marcador de peso molecular 1 Kb plus; C- Controle negativo.	50
Figura 10 - Eletroforese em gel de agarose dos produtos da PCR com os primers CR1-Mix realizado com produtos de RCAs das amostras de coqueiro. M- Marcador de peso molecular 1 Kb plus; C- Controle negativo.	51
Figura 11 - Eletroforese em gel de agarose dos amplicons obtidos da PCRM com os primers CR1-Mix e purificados por eluição. M- Marcador de peso molecular 1 Kb plus; C- Controle negativo; C+ controle positivo.	52
Figura 12 - Logotipo das sequências da CR-M alinhadas dos isolados de CCAV, em verde, a localização da região rica em CG na região comum maior. Fonte: Autora.	54
Figura 13 - Estrutura secundária do hair-pin dos componentes genômicos do CCAV, DNA-1, DNA-2, DNA-3, DNA-4, DNA5-, DNA6, DNA7 e DNA-7_1. Fonte: Autora.	55
Figura 14 - Alinhamento múltiplo das sequências de nucleotídeos de cada componente genômico completo separadamente utilizando o programa MUSCLE. Em preto, as sequências e o consenso entre elas. Em verde, o perfil de identidade no decorrer do alinhamento. Fonte: Autora.	57
Figura 15 - Porcentagens de identidade da sequência de nucleotídeos dos DNAs genômicos do CCAV calculadas com o programa de demarcação de sequência (SDT). Fonte: Autora.	58
Figura 16 - Representação da caracterização molecular dos componentes genômicos do CCAV, em amarelo a representação da ORF dos componentes, em vermelho a região comum maior (CR-M) e em verde, a posição do stem loop (AAGTATTAC) Fonte: Autora.	62
Figura 17 - Identificação a partir do software Pfam de motif relacionado a Rep nas sequências de proteína codificadas pelo componente 7 dos isolados 4, 12 e 22 do CCAV. Em vermelho, a região que o software apontou como motifs helicase e Viral_Rep.	63

Figura 18 - Representação gráfica da variação em aminoácidos dos motifs RCR I, II e III e domínios helicase (Walker A, B e C) em representantes dos gêneros Babuvirus e Nanovirus e dos isolados de CCAV.	64
Figura 19 - Ilustração da localização dos seis motifs conservados na Rep de CCAV, com os três motivos RCR (replicação círculos rolante) e os motivos de helicase Waker A, B e C em marrom.	64
Figura 20 - Árvore filogenética de máxima verossimilhança construída a partir de um alinhamento de sequências codantes do componente Rep de vírus pertencentes aos gêneros Cocosatellite, Muscarsatellite, Clostunsatellite e Subclovsatellite (família Alphasatellitidae), Babuvirus e e Nanovirus (família Nanoviridae). A árvore foi enraizada por midpoint. Os círculos pretos mostram os ramos com suporte maior que 0,95.....	65

LISTA DE TABELAS

Tabela 1 - Área colhida (em hectares), produção (frutos), valor da produção (R\$) de coco-da-baía (Cocos nucifera L.), por regiões do Brasil.....	21
Tabela 2 - Origem e funções das proteínas codificadas pelos componentes genômicos dos nanovírus e babuvírus da família Nanoviridae.....	24
Tabela 3 - Espécies de vírus pertencentes ao gênero Nanovirus reconhecidas pelo ICTV.	25
Tabela 4 - Espécies de Babuvirus reconhecidas pelo ICTV.	29
Tabela 5 - Sequenciadores da Oxford Nanopore Technologies.	31
Tabela 6 - Sequência dos primers utilizados para amplificação do componente Rep.	37
Tabela 7 - Sequência dos primers utilizados na PCR multiplex para obtenção dos componentes genômicos do CCAV.	41
Tabela 8 - Vírus pertencentes aos gêneros Cocosatellite, Muscarsatellite, Clostunsatellite e Subclovsatellite, Babuvirus, e Nanovirus e números de acessos no GenBank dos componentes Rep utilizados nesse trabalho.....	42
Tabela 9 - Caracterização dos componentes genômicos do coconut crown atrophy vírus. Fonte: Autora.	53

LISTA DE QUADROS

Quadro 1 - Relação de amostras de coqueiro. Número referente a cada amostra, local de coleta código da amostra, e presença de sintoma.....	33
Quadro 2 - Protocolo de miniprep (KOTCHONI et al., 2003).....	38

LISTA DE SIGLAS E ABREVIATURAS

µg- micrograma

µL- microlitro

µM- micromolar

aa- aminoácido

BLAST- basic local alignment search tool

CTAB- brometo de cetiltrimetilamônio

DF- Distrito Federal

DNA- ácido desoxirribonucleico

dNTPs- desoxirribonucleotídeos fosfatados

ssDNA- DNA de fita simples

EDTA- ácido etilenodiamino tetra-acético

ICTV- international committee on taxonomy of viruses

Kb- quilo base

KCl- cloreto de potássio

LPP3- laboratório de interação planta-praga III

M- Molar

mM- milimolar

MgCl₂- cloreto de Magnésio

NCBI- National center for biotechnology information

NEB- New england biolabs inc.

ORF- open reading frame

pb- pares de base

PCR- reação em cadeia da polimerase

RCA- amplificação por círculo rolante

Rep- proteína de replicação viral

NGS- sequenciamento de nova geração

LISTA DE SIMBOLOS

°C Grau Celsius

® Registrado

μ Micro

V Volt

% Porcentagem

Sumário

Introdução.....	17
Objetivo	20
Objetivos Específicos.....	20
Referencial Teórico	20
Produção de coco no Brasil	20
Atrofia da coroa do coqueiro (ACC).....	21
Família <i>Nanoviridae</i>	23
Gênero <i>Nanovirus</i>	25
Organização genômica e replicação dos nanovírus.....	26
Os componentes genômicos dos nanovírus não necessariamente ocorrem simultaneamente para expressão de sintomas	28
Gênero <i>Babuvirus</i>	28
Critério de demarcação de espécies na família <i>Nanoviridae</i>	29
Sequenciamento alto desempenho (<i>High throughput sequencing</i> - HTS).....	29
Metodologia.....	33
Obtenção dos isolados	33
Extração de DNA e Amplificação por círculo rolante (RCA)	36
Detecção do componente genômico Rep e purificação dos produtos de PCR	36
Clonagem e sequenciamento do componente Rep.....	37
Análise das sequências Rep obtidas dos Isolados 4 e 12	39
Desenho de <i>primers</i> e amplificação de componentes genômicos em reação multiplex	40
Sequenciamento Nanopore e montagem dos componentes genômicos	41
Análise filogenética dos componentes Rep do CCAV	42
Cálculo de identidades dos componentes dos diferentes isolados	44
Identificação de <i>motifs</i> conservados da Rep	44
Resultados e discussão.....	44
Detecção, clonagem e análise molecular do componente genômico Rep dos isolados 4 e 12	44
Sequenciamento Nanopore aplicado a caracterização do CCAV	52
Caracterização molecular dos genomas a partir do sequenciamento Nanopore	53
Análise filogenética das sequências Reps dos isolados do CCAV	65
Perspectivas futuras	67
Anexo I	68
Referências	70

Introdução

No Brasil, o cultivo de coqueiro (*Cocos nucifera* L.) é um dos destaques da produção agrícola nacional. Aproximadamente 200 mil hectares são direcionados ao plantio dessa cultura com uma produção de 1,95 bilhões de frutos (IBGE, 2020), fazendo com que o país ocupe a quinta posição dentre os maiores produtores de coco, com participação de 4,5% na produção mundial (MAPA, 2021). Os principais estados produtores em escala decrescente são Bahia, Ceará, Pará, Sergipe, Espírito Santo e Pernambuco, que desempenham grande atuação no faturamento da exportação anual estimado em R\$ 1 bilhão de reais a partir da produção de coco e produtos derivados da matéria prima (FERRAZ et al., 2020).

Assim como acontece em outras culturas, o coco também é suscetível a pragas e problemas fitossanitários que podem prejudicar seu cultivo (FERREIRA; WARWICK; SIQUEIRA, 2018) como o "lethal yellowing" ou amarelecimento letal (fitoplasmas letais do grupo amarelecimento 16Sr IV), resinose (*Thielaviopsis paradoxa*), ácaro da palmeira vermelha (*Raoiella indica* Hirs), ácaro do coco (*Aceria guerreronis* Keifer), doença do anel vermelho (*Bursaphelenchus cocophilus*) (GAMA; ROBERTO; FARIAS, 2021), dentre outras. No Brasil, uma doença conhecida como "atrofia letal da coroa do coqueiro" (ALCC) foi registrada em Pernambuco no ano de 2012 afetando a produtividade de coqueiros no estado. Até onde se relata, a doença também já acometeu coqueiros da Bahia, Pará, Paraíba, Roraima, Sergipe e Amazonas (FERRAZ et al., 2020).

Alguns autores retiraram o termo "letal" do nome da doença (BOARI et al., 2019; KAUFFMANN et al., 2019). Assim, ela também é conhecida como "atrofia da coroa do coqueiro" (ACC) tendo em vista que não se observou a morte de coqueiros em nove plantios distribuídos em quatro estados visitados. Além disso, é comum observar a recuperação de coqueiros com ACC ao longo do tempo. Sendo assim, neste trabalho, vamos adotar o nome "atrofia da coroa do coqueiro" (ACC), como descrito na literatura mais recente.

A ACC tem como sintomas principais a perda dos frutos, necrose, assim como o encurtamento foliar, somado a isso, a grande problemática em torno da doença está na falta de informações relacionadas a sua etiologia bem como possíveis vetores transmissores. Ferraz *et al.* (2020) levantaram a hipótese sobre a associação da doença a *Fusarium* sp, que foi isolado em 84/130 coqueiros jovens apresentando a doença. No entanto, não houve comprovação da patogenicidade deste fungo ao coqueiro. Ainda no mesmo trabalho, os autores ressaltam

algumas particularidades da ACC, como a independência de fatores (genética, idade, solo, sistema de plantio, altitude e extensão de área), longevidade variável das plantas, propagação também variável (possibilidade da ocorrência de coqueiros sadios próximos a coqueiros com a doença) e sobre a recuperação, ocorrendo êxito e não êxito sob tratamentos agroquímicos, uso de sulfato de zinco e na paralização da recuperação.

BOARI e colaboradores (2019), estudando coqueiros com ACC observaram DNA de cerca de 1.000 pb após a restrição enzimática do DNA amplificado por meio do RCA, tamanho compatível com o genoma da família *Nanoviridae*. Assim, por meio de um sequenciamento de nova geração de amostras de DNA amplificado via RCA descobriram a presença de um DNA-REP semelhante a DNA-REP associados a membros da família *Nanoviridae*. A partir disso, o mesmo grupo desenvolveu um protocolo (RCA-PCR) para detecção desse componente genômico em amostras de coqueiro com ACC provenientes do Pará, Paraíba, Pernambuco e Bahia utilizando o par de *primer* NanoRep-F e R uma vez que o par Babu R e F não permitiram a detecção do DNA-REP nas amostras provenientes de Capanema-PA e estados do Nordeste (KAUFFMANN et al., 2019).

A família *Nanoviridae* é composta por um grupo de vírus multipartidos e compreende dois gêneros reconhecidos oficialmente: *Nanovirus* e *Babuvirus*. Os genomas dos membros desses gêneros são divididos em componentes genômicos circulares de DNA de fita simples (DNA-R, C, S, M, N, U) que possuem de 920-1.100 nucleotídeos e que codificam de cinco a sete proteínas e são encapsulados individualmente em estruturas icosaédricas de simetria T=1 (BURNS; HARDING; DALE, 1995; SELVARAJAN, 2019; VETTEN, 2008).

O DNA-R codifica a proteína iniciadora de replicação Rep, também conhecida como *master Rep* (M-Rep), nome dado pela sua participação na replicação de todos os componentes genômicos do vírus (TIMCHENKO et al., 1999). Além do DNA-R, alguns nanovírus estão associados a moléculas do tipo *alphasatellite*, que codificam uma proteína *Rep-like*. Entretanto, essas moléculas *alphasatellites* não são capazes de replicar os componentes genômicos dos nanovírus (TIMCHENKO et al. 1999), podendo replicar apenas os seus próprios genomas (HORSER; HARDING; DALE, 2001). O DNA-C, S, N e M codificam respectivamente para proteína “cell-cycle link” (*Clink*) responsável pela interação com proteínas tipo retinoblastoma interferindo no ciclo celular da planta (LAGEIX et al., 2007); proteína capsidial, proteína de movimento, proteína “nuclear-shuttle protein” responsável pelo transporte intracelular do ssDNA núcleo ao citoplasma do DNA viral sendo indispensável também na transmissão do

vírus por pulgões (GRIGORAS et al., 2018) e a proteína de movimento que realiza o transporte intracelular do complexo NSP-ssDNA (MOROZOV & SOLOVYEV, 2020). Já a função dos DNA-Uss não está completamente esclarecidas na literatura (SELVARAJAN, 2019).

Nesse contexto de doenças em plantas ocasionadas por patógenos, é de extrema importância a detecção e identificação confiável de qual/ quais microrganismos se trata para traçar a melhor estratégia para o controle da doença (CHALUPOWICZ et al., 2019). O avanço em tecnologias com base no sequenciamento de alto desempenho (*High-Throughput Sequencing Technologies* – HTS) junto a análises de bioinformática trouxeram gigantescos avanços para virologia vegetal, tornando possível não só a identificação de novos vírus como a sua caracterização (FAJARDO et al., 2017).

O sequenciamento nanopore é considerado a quarta geração do HTS. Essa tecnologia traz consigo o conceito de fácil portabilidade e manuseio. Ela se baseia em um poro de proteína que funciona como um biosensor anexado a um polímero eletricamente resistente, com isso, à medida que o DNA atravessa o "nanoporo" são emitidas perturbações elétricas específicas a cada nucleotídeo, e como resultado, é possível obter a leitura de longas sequências em tempo real (WANG et al., 2020).

Nessa dissertação, é descrita e validada uma metodologia para o uso do sequenciamento nanopore a partir de *primers* com potencial de acesso a diversidade genética em uma abordagem *multiplex*. Essa metodologia permite detectar e realizar a caracterização molecular de um vírus multipartido identificado em amostras de coqueiros do Brasil apresentando a atrofia da coroa do coqueiro. A estratégia metodológica dispõe de mais rapidez na detecção e caracterização bem como um menor custo na obtenção dos resultados para virologia vegetal. A pesquisa tem como base resultados prévios obtidos pela Dra. Alessandra de Jesus Boari, pesquisadora da Embrapa Amazônia oriental e sua equipe (BOARI et al., 2019; KAUFFMANN et al., 2019).

Objetivo

Descrever e validar metodologia de identificação de um vírus multipartido associado a atrofia da coroa do coqueiro (ACC) a partir de *primers* permitindo acesso a diversidade molecular viral fazendo uso do sequenciamento nanopore.

Objetivos Específicos

- Detectar o componente Rep nas amostras de coco, clonagem e sequenciamento;
- Desenhar *primers* específicos para acessar a diversidade molecular (obtenção de todos os componentes genômicos) do vírus em amostras coletadas no Pará, Paraíba, Roraima Bahia e Pernambuco;
- Amplificar por RCA-PCR as amostras de coqueiro com os primers específicos;
- Realizar o sequenciamento nanopore das amostras positivas para o vírus, identificando os 7 componentes genômicos;
- Caracterizar molecularmente os componentes genômicos obtidos;
- Analisar filogeneticamente as sequências (nucleotídeos) genômicas do componente DNA-REP

Referencial Teórico

Produção de coco no Brasil

O coqueiro (*Cocos nucifera* L.) é uma palmeira tropical que pode atingir até 30 metros com folhas de até 3 metros de comprimento pertencente à família Arecaceae (MAGALHÃES, NEY et al., 2017). Evidências históricas e filogenéticas apontam que o coqueiro se originou no sudeste asiático. No Brasil, foi introduzido através da Bahia em 1553 por portugueses, por isso é conhecido tipicamente como coco-da-baía. Depois propagou-se por toda a costa brasileira por dispersão natural (SIQUEIRA; ARAGÃO; & TUPINAMBÁ, 2002).

Os tipos de coqueiros se dividem em três variedades principais, sendo eles, o coqueiro híbrido, coqueiro gigante e coqueiro anão (FONTES; FERREIRA, 2006). A atratividade para os produtores da cultura do coco está na infinidade de produtos e subprodutos que podem ser obtidos a partir do coqueiro. O aproveitamento da planta é praticamente integral, tendo empregabilidade para fins alimentícios, agroindustriais, medicinais e artesanais (MACHADO et al., 2020).

No Brasil, o coqueiro é uma cultura de grande importância econômica, principalmente na comercialização do coco verde ou coco seco (LIMA, 2013). Dentre os produtos comercializáveis, a água de coco tem destaque por se tratar de uma bebida natural rica em vitaminas, minerais proteínas e açúcares com baixo teor calórico (ARAGAO; ISBERNER; CRUZ, 2001). Segundo o Ministério da Agricultura Pecuária e Abastecimento (MAPA, 2021), o país é o principal produtor mundial de água de coco e quinto maior produtor mundial de frutos do coqueiro, segundo dados do (IBGE, 2020) a produção em 2020 foi de 1.951.000 frutos apresentando a região nordeste como principal produtora (Tabela 1).

Tabela 1 - Área colhida (em hectares), produção (frutos), valor da produção (R\$) de coco-da-baía (*Cocos nucifera* L.), por regiões do Brasil.

País/Regiões	Área colhida (hectares)	Produção (milhões de frutos)	Valor da produção (reais)
			Ano 2020 2018
Brasil	223.057	1.951	972.962,00
Norte	33.698	361	122.320,00
Nordeste	169.763	1.269	674.867,00
Sudeste	17.387	297	156.052,00
Sul	522	5	2.279,00
Centro-Oeste	1.686	19	17.445,00

Fonte: Dados do IBGE.

A região Nordeste é a maior produtora de coco do país, exibe um percentual de 76% de toda área colhida em hectares e 65% da produção nacional em frutos. Essa posição se mantém desde a introdução dos plantios no país, que tiveram ocorrência no litoral da região. Devido a pandemia de SARS-CoV-2 ocorreu uma queda na demanda, possivelmente devido ao isolamento social, entretanto as perspectivas para o primeiro semestre de 2023 é de aumento de procura, isso considerando a aptidão de resposta do setor produtivo e o poder de compra dos consumidores do produto (BRAINER; XIMENES, 2020).

Atrofia da coroa do coqueiro (ACC)

A preocupação da comunidade nucícola internacional está há anos voltada a uma doença denominada "*lethal yellowing*" ou em português "amarelecimento letal" do coqueiro, que vem se disseminando em plantações de coqueiro no mundo. No Brasil, ainda não se tem relatos sobre essa doença. Entretanto, os coqueiros do país não estão isentos de patologias que ameaçam os 300 mil empregos ofertados e faturamento anual de R\$ 1bilhão (FERRAZ et al., 2020).

A "resinose" causa o afinamento do tronco, lesões necróticas na raiz, baixa frequência na formação de novas folhas, as já existente ficam amareladas e frágeis, e em casos extremos, cachos secos e com coloração escura, que comprometem completamente a produção (FERREIRA, 1998). O "ácaro vermelho das palmeiras" possui o potencial de reduzir a produção de coco em até 70% (DINIZ, 2009), e uma nova doença observada em coqueiros safreiros (coqueiro anão, gigante e híbrido) vem levantando alertas em alguns estados do Brasil como Pernambuco, Amazonas, Pará, Paraíba, Bahia, Sergipe e Roraima (FERRAZ et al., 2020).

O nome "atrofia letal da coroa do coqueiro" foi dado devido aos sintomas que caracterizam a doença, ocorrendo inicialmente em folhas jovens onde os folíolos se tornam necróticos e torcidos em suas extremidades, ocasionando a redução do seu comprimento, Figura 1. Como resultado, a atrofia de todas as folhas da planta, e em casos extremos, ocorre a queda das copas remanescentes e morte do coqueiro (FERRAZ, 2012).

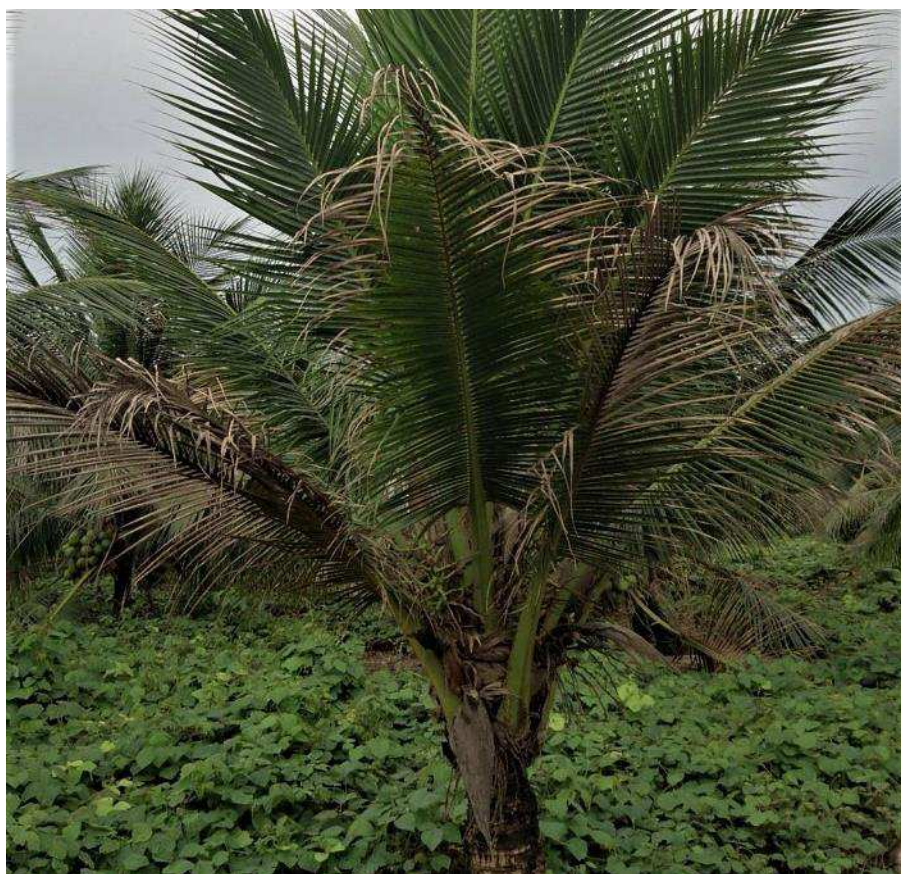


Figura 1 - Coqueiro apresentando a doença atrofia da coroa do coqueiro. Fonte: Boari, dados não publicados.

Inicialmente, a ACC em coqueiros foi apontada como uma deficiência de boro. Essa hipótese foi descartada a partir da análise de tecidos foliares de coqueiros sintomáticos

(FERRAZ, 2012), abrindo espaço para a suspeita da deficiência em zinco, cofator importante na síntese do triptofano que origina o ácido indolacético, responsável pelo crescimento das plantas (MENGEL, KONRAD; KIRKBY, 2012). Nessa linha de pensamento, Ferraz e colaboradores (2020) levantam a hipótese de a deficiência em zinco ocasionar a infecção oportunista do fungo *Fusarium* sp, encontrado em 65% (130 plantas) sintomáticas dos estados do Pernambuco e da Paraíba. Além disso, há dados publicados, disponibilizados em parceria pela Dra. Alessandra Jesus de Boari, fitopatologista da Embrapa Amazônia Oriental e colaboradores da presença de um vírus com características similares a um vírus da família *Nanoviridae* em amostras de coqueiros sintomáticos para ACC (BOARI et al., 2019).

Família *Nanoviridae*

Os membros da família *Nanoviridae* surgiram como importantes patógenos que naturalmente infectam plantas monocotiledôneas e dicotiledôneas (KATUL et al., 1998; MAKKOUK; KUMARI, 2009; SAUNDERS; STANLEY, 1999), atingindo severamente culturas em diversas regiões no mundo como o oeste Asiático, norte da África, Espanha, Japão, China e Austrália (ORTIZ et al., 2006); (SANO et al., 1998); (CHU; HELMS, 1989). Dividida entre os gêneros *Nanovirus* e *Babuvirus*, a família possui atualmente quatorze espécies reconhecidas pelo comitê internacional de taxonomia de vírus, sendo onze pertencentes ao gênero *Nanovirus* e três referentes ao *Babuvirus*, (INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES ICTV, 2021).

O genoma dos membros da família é multipartido, sendo composto por 6 ou 8 componentes de DNA circular de fita simples (ssDNA) (VETTEN et al., 2005). São encapsulados separadamente em partículas de estrutura icosaédrica T=1 com diâmetro que varia entre 17-20nm (ABRAHAM;VARRELMANN & VETTEN., 2012) (Figura 2).

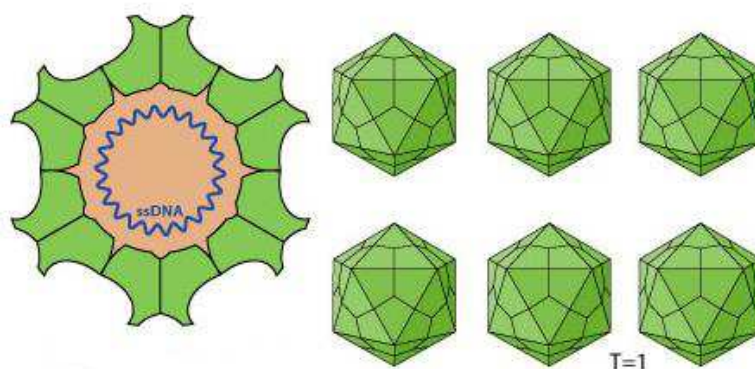


Figura 2 - Estrutura do capsídeo dos nanovírus em forma icosaédrica contendo o ssDNA em seu interior. Fonte: (VIRALZONE, 2009; https://viralzone.expasy.org/125?outline=all_by_species).

Os componentes do genoma ou DNAs genômicos, são nomeados de acordo com a função específica de cada proteína que ele codifica dentre eles, o DNA-R, DNA-S, DNA-C, DNA-M e DNA-N são comuns entre os nanovírus e os babuvírus, enquanto DNA-U1, DNA-U2 e DNA-4 são específicos dos nanovírus e o DNA-U3 dos babuvírus (Tabela 2)(ABRAHAM;VARRELMANN & VETTEN., 2012; INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES ICTV, 2021; SELVARAJAN, 2019)

Tabela 2 - Origem e funções das proteínas codificadas pelos componentes genômicos dos nanovírus e babuvírus da família *Nanoviridae*.

Componente genômico	Proteína codificada	Identificado		Função da proteína
		Nanovírus	Babuvírus	
DNA-R	M-Rep	+	+	Iniciadora de replicação
DNA-S	CP	+	+	Estrutura do capsídeo
DNA-S	Clink	+	+	Ciclo celular do hospedeiro (interação com proteínas tipo retinoblastoma)
DNA-M	MP	+	+	Movimento célula a célula
DNA-N	NSP	+	+	Transporte intracelular do ssDNA
DNA-U1	U1	+	-	Desconhecida
DNA-U2	U2	+	-	Desconhecida
DNA-U3	U3	-	+	Desconhecida
DNA-U4	U4	+	-	Desconhecida

Fonte: Autora com base nas informações do ICTV, 2021.

Gênero *Nanovirus*

Os nanovírus infectam predominantemente plantas dicotiledôneas leguminosas (GRIGORAS et al., 2018; TIMCHENKO et al., 1999), que têm um papel primordial na cadeia de consumo alimentar humano e animal, e também têm importante utilização como adubos verdes (GAAFAR; ZIEBELL, 2016). A sintomatologia provocada pelos nanovírus é variada, podendo ocorrer necrose dos tecidos foliares, amadurecimento precoce, nanismo e até mesmo a morte precoce das plantas hospedeiras (ABRAHAM;VARRELMANN & VETTEN., 2012; GAAFAR; ZIEBELL, 2020, 2016).

Até o momento, não há nenhum relato de transmissão de nanovírus por meios mecânicos ou por sementes, devido, provavelmente, a sua restrição ao floema das plantas hospedeiras. Desse modo, dependem de vetores para a transmissão planta a planta. Na natureza, são transmitidos por uma variedade de pulgões, tendo como espécies hemípteras de destaque *Aphis craccivora*, *Acyrtosiphon pisum* e *A. Fabae* (ABRAHAM;VARRELMANN & VETTEN., 2012; GRIGORAS et al., 2009). A transmissão do vírus através do vetor ocorre de forma circulativa persistente e não propagativa. Isso significa que os pulgões precisam se alimentar efetivamente da planta infectada para adquirir as partículas virais e as partículas necessitam passar do intestino para a hemolinfa, e posteriormente, para as glândulas salivares para que os pulgões virulíferos possam injetar as partículas virais quando se alimentarem em novas plantas (ABRAHAM;VARRELMANN & VETTEN., 2012). Sendo não propagativa implica que os nanovírus não se replicam em seus insetos vetores (GAAFAR; ZIEBELL, 2020).

Atualmente, nove espécies pertencentes ao gênero *Nanovirus* são reconhecidas pelo Comitê Internacional de Taxonomia de Vírus (Tabela 3).

Tabela 3 - Espécies de vírus pertencentes ao gênero *Nanovirus* reconhecidas pelo ICTV.

Espécies de Nanovírus	Referências
<i>Black medic leaf roll virus</i>	(GRIGORAS et al., 2014a)
<i>Cow vetch latent virus</i>	(GALLET et al., 2018)
<i>Faba bean necrotic stunt virus</i>	(GRIGORAS et al., 2009)
<i>Faba bean necrotic yellows virus</i>	(KATUL et al., 1997); (KATUL et al., 1998b)
<i>Faba bean yellow leaf virus</i>	(ABRAHAM;VARRELMANN & VETTEN., 2012)
<i>Milk vetch dwarf virus</i>	(SANO et al., 1998)
<i>Parsley severe stunt associated virus</i>	(VETTEN et al., 2019)
<i>Pea necrotic yellow dwarf virus</i>	(GRIGORAS et al., 2010)
<i>Pea yellow stunt virus</i>	(GRIGORAS et al., 2014b)
<i>Sophora yellow stunt virus</i>	Não publicado

Organização genômica e replicação dos nanovírus

Anteriormente conhecidos como “circovírus de plantas”, em 1998 o termo nanovírus (PRINGLE, 1998) foi atribuído a esse grupo de vírus devido ao formato de seus vírions e componentes do genoma, que é formado por um DNA de fita simples.

Membros do gênero *Nanovirus* são constituídos por oito componentes genômicos de ssDNA com cerca de 0,9 a 1,0kb de tamanho (VETTEN, 2008).

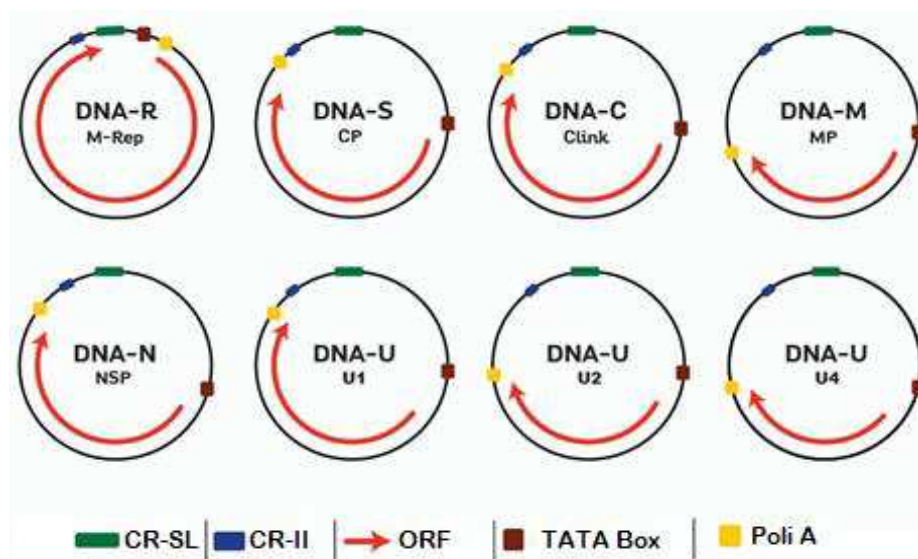


Figura 3 - Componentes genômicos dos nanovírus DNA-R, S, C, M, N, U1, U2 e U4 com suas respectivas proteínas codificadas; CR-SL, região comum I (região formadora do Steam Loop, M-rep reconhecem a origem de replicação); CR-II segunda região de sequências conservadas; ORF, open Reading frame (início de região codificadora de proteínas ou RNAs funcionais para sequência de DNA); TATA Box, sequência promotora reconhecida pelos fatores de transcrição; Poli A, sinal de poliadenilação. **Fonte:** Figura elaborada pela autora de acordo com informações sobre o genoma do nanovírus faba bean necrotic yellows virus (FBNYV)

O DNA-R, codificador da proteína M-Rep, a rep principal é necessária para iniciar a replicação por círculo rolante de todos DNAs genômicos (TIMCHENKO et al., 1999). O DNA-S, contém o gene responsável pela constituição estrutural do vírus, sendo codificador da proteína do capsídeo - (CP) (CHU et al., 1993). O DNA-C, codificando a proteína “cell cycle link” (Clink) é responsável pela modulação do ciclo celular estimulando a replicação viral dos nanovírus nos hospedeiros (ARONSON et al., 2000). Experimentos *in vivo* demonstraram a capacidade da *Clink* em se ligar a proteína do tipo retinoblastoma induzindo células inativas a

progressão do ciclo celular (LAGEIX et al., 2007). O DNA-M, codifica a proteína responsável pelo movimento célula a célula “MP”. O DNA-N é codificador da “nuclear shuttle protein” (NSP) que é responsável pelo transporte do DNA viral para fora do núcleo (WANITCHAKORN, R., HAFNER, G. J., HARDING, R. M., & DALE, 2000). Embora o capsídeo viral seja indispensável na transmissão de nanovírus por pulgões, foi demonstrado por GRIGORAS e colaboradores (2018) que a NSP age como fator auxiliar na transmissão pelo vetor, mas não participa diretamente na formação do vírion ou nos sintomas da doença.

Ainda segundo GRIGORAS e colaboradores (2018), é pertinente destacar que não se sabe se proteínas com função ainda não estabelecidas podem interferir na interação vírus vetor, visto que se observou redução na eficiência da transmissão quando o componente DNA-U1 está ausente. As proteínas codificadas pelo DNA-U1, DNA-U2 e DNA-U4 não apresentam funções claras descritas na literatura. Entretanto, segundo (GRIGORAS et al., 2018) e (GAAFAR; ZIEBELL, 2020) as proteínas codificadas por DNA-U1 e DNA-U2 resultaram em uma redução na severidade dos sintomas, mas a transmissão ainda era possível, enquanto a eliminação do DNA-U4 não apresentou alteração na infectividade ou sintomatologia.

Os nanovírus fazem uso do mecanismo de círculo rolante (TIMCHENKO et al., 1999) para sua replicação, onde o genoma circular de ssDNA serve como molde para a síntese de várias cópias do material genético (ZERBINI et al., 2017).

Os componentes genômicos dos nanovírus possuem uma ORF com transcrição unidirecional (fita positiva,) capaz de codificar uma única proteína. Cada região codificadora é antecedida por uma sequência promotora TATA e seguida na região 3’ da ORF por uma sequência AATAAA que funciona como um sinal de poliadenilação. Entretanto, para o DNA-R o sinal de poliadenilação se encontra entre TATA e a ORF (INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES ICTV, 2021; THOMAS et al., 2021).

A proteína M-rep é responsável pela replicação de todos os componentes virais dos nanovírus e essas reações bioquímicas envolvem duas regiões dentro do genoma viral (SELVARAJAN, 2019). A região comum I (CR-I ou CR-SL) possui sequências com repetições invertidas que formam uma estrutura *stem-loop*. A sequência formadora da região da “alça” contém o nonanucleotídeo TAGTATTAC, altamente conservado nos nanovírus e que se acredita ser o local de origem de replicação. Dessa forma, as proteínas M-Rep reconhecem a origem de replicação viral e clivam em uma sequência conservada desencadeando o início da síntese do DNA viral no sentido positivo que depois será seguida pela replicação em círculo

rolante realizado por uma DNA polimerase celular (TIMCHENKO et al., 1999). A CR-I possui alta identidade entre os componentes de um nanovírus pois, possíveis mudanças nessa região podem afetar negativamente a eficiência replicativa. Enquanto a região comum II, que também possui sequências semelhantes, tem 20 nucleotídeos e está localizada na extremidade 5' da CR-I possuindo identidade entre os componentes genômicos dos nanovírus que variam de 75-100%, isso levando em consideração o faba bean necrotic stunt virus (FBNSV) (KATUL et al., 1997).

Os componentes genômicos dos nanovírus não necessariamente ocorrem simultaneamente para expressão de sintomas

Um aspecto que chama a atenção da virologia se tratando de vírus multipartidos é como esses vírus conseguem infectar eficientemente as células individualmente com todos os componentes genômicos, tendo em vista que cada componente tem sua encapsulação individual. Nesse sentido, quais componentes são essenciais para replicação de um nanovírus? TIMCHENKO; KATUL; ARONSON (2006) demonstraram pela primeira vez em um estudo sobre a infectividade dos componentes genômicos de faba bean necrotic yellow virus (FBNYV) que cinco componentes, o DNA-R, DNA-M, DNA-S, DNA-U1 e DNA-U2 foram o suficiente para induzir sintomas da doença em plantas de *Vicia faba*.

Buscando respostas para a baixa probabilidade de uma célula conseguir comportar todos os componentes genômicos de um nanovírus, (SICARD et al., 2019) demonstraram por imunofluorescência que os componentes de faba bean necrotic stunt virus (FBNSV) não infectam necessariamente em uma mesma célula, mas podem estar presentes em diferentes células do tecido da planta. Os diferentes componentes se complementam em células vizinhas propiciando uma infecção multicelular.

Gênero *Babuvirus*

Diferentemente dos nanovírus, membros do gênero *Babuvirus* infectam predominantemente plantas monocotiledôneas. Os pulgões *Micromyzus kalimpongensis* (cardamom bushy dwarf virus), *Pentalonia nigronervosa* (abaca bunchy top virus) e *P. caladii* (banana bunchy top virus) são vetores naturais dos três vírus pertencentes ao gênero

(INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES ICTV, 2021). O genoma de ssDNA circular é segmentado em 6 componentes genômicos que possuem 1-1.1kb, DNA-R, DNA-S, DNA-C, DNA-M, DNA-N com as mesmas funções já descritas acima para nanovírus, e DNA-U3, que é restrito aos babuvírus e não tem sua função completamente conhecida (SELVARAJAN; BALASUBRAMANIAN, 2014). Na Tabela 4, são evidenciadas as espécies associadas ao gênero *Babuvirus* segundo o ICTV (2021).

Tabela 4 - Espécies de *Babuvirus* reconhecidas pelo ICTV.

Espécies de Babuvirus	Referências
<i>Abaca bunchy top virus</i>	(SHARMAN et al., 2008)
<i>Banana bunchy top virus</i>	(BURNS; HARDING; DALE, 1995)
<i>Cardamom bushy dwarf virus</i>	(MANDAL et al., 2013)

Fonte: Autora com base nas informações do ICTV, 2021.

Critério de demarcação de espécies na família *Nanoviridae*

Segundo o comitê internacional de taxonomia dos vírus (INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES ICTV, 2021) quatro critérios são propostos para o encaixe de novas espécies dentro dos gêneros da família *Nanoviridae*: i) A identidade de nucleotídeos global maior que 80% combinando todos os componentes; ii) Diferença nos hospedeiros; iii) Diferença na identidade e variedade de espécies de pulgões vetores (ou seja, vetores em potencial) e iv) reações de anticorpos a membros de espécies divergentes.

O coconut foliar decay virus (CFDV) tido anteriormente como um possível nanovírus, apresenta 7 componentes com identidade a *alphasatellites*, DNA-S1 e DNA-S2 que codificam proteínas com similaridade a CP, um componente genômico ainda menor, DNA- gamma que os autores especulam ser o precursor da proteína de movimento do CFDV. Esse vírus parece representar um novo gênero e uma nova família de vírus (GRONENBORN et al., 2018).

Sequenciamento alto desempenho (*High throughput sequencing* - HTS)

Em 1977 as primeiras tecnologias básicas de sequenciamento de DNA surgiam. O método de sequenciamento Maxam-Gilbert por degradação (MAXAM; GILBERT, 1977) e o método de síntese de Sanger ou também conhecido como Didesoxinucleotídeos Sanger

(SANGER; NICKLENX; COULSON, 1977). O método Maxam- Gilbert se baseia na modificação química seguida pela clivagem estrutural do DNA em locais próximos aos nucleotídeos modificados. O sequenciamento Sanger faz uso de nucleotídeos que se encontram nas cadeias terminais (didesoxinucleotídeos) que não apresentam o grupo 3'- OH, com isso a DNA polimerase fica impossibilitada de formar ligações fosfodiéster culminando na terminação da cadeia de DNA (SLATKO; GARDNER; & AUSUBEL, 2018).

As técnicas refletiam seus resultados através de gel poliacrilamida realizando a separação do DNA por tamanho (GUŽVIĆ, 2013) Entretanto, embora o método químico desenvolvido por Maxam-Gilbert tenha sofrido modificações para eliminação de agentes químicos, o método de Sanger foi posteriormente automatizado se tornando o método padrão para sequenciamento de DNA (SLATKO; GARDNER; & AUSUBEL, 2018) e, alcançando grandes feitos como a conclusão do projeto genoma humano 1977-2004 (MARGULIES et al., 2005).

Com o passar dos anos, algumas limitações do método Sanger como o elevado tempo para obtenção de resultados, baixo rendimento de leitura e alto custo, passaram a se tornar um obstáculo no avanço do sequenciamento (KCHOUK; GIBRAT; ELLOUMI, 2017). Com isso, após trinta anos de seu desenvolvimento, novas tecnologias de sequenciamento começaram a ser comercializadas, e esses métodos passaram a ser chamados de tecnologias de sequenciamento de nova geração (*Next Generation Sequencing* - NGS) (GOODWIN; MCPHERSON; MCCOMBIE, 2016).

Os métodos de sequenciamento de segunda geração surgiram como uma melhoria na eficiência do sequenciamento de genomas maiores, visando um menor custo no processo. As principais tecnologias de sequenciamento de segunda geração são: 454 pirosequenciamento (Roche), Ion Torrent Tecnologia (Life technologies), (ILLUMINA). A plataforma Illumina em especial, até 2016 era responsável por cerca de 90% do volume de dados sequenciados no mundo todo, isto devido ao seu baixo custo, alto volume de dados produzidos e elevada precisão (GOODWIN; MCPHERSON; MCCOMBIE, 2016).

Ao contrário do sequenciamento de segunda geração, o de terceira visa o sequenciamento de longas sequências de DNA (e RNA), sendo possível a marcação fluorescente de um único nucleotídeo para identificação de sequências longas (~ 175kb), sem a necessidade de pré-amplificação dos fragmentos a serem sequenciados (GOODWIN;

MCPHERSON; MCCOMBIE, 2016). O primeiro sequenciamento em tempo real de uma única molécula (*Single-Molecule Real Time*) foi lançado em 2010 pela Pacific Biosciences (VAN DIJK et al., 2014).

A considerada quarta geração de sequenciadores de nova geração, se baseia na passagem do DNA através de poros de pequeno diâmetro com capacidade de identificar nucleotídeo a nucleotídeo através de diferentes correntes geradas por um detector ligado à medida em que o DNA está sendo sequenciado (LIU et al., 2016). Em 2005, a Oxford Nanopore Technologies foi fundada com o intuito de desenvolver tecnologias de sequenciamento baseadas na detecção de uma molécula única fazendo uso da tecnologia de nanoporos. Seu primeiro sequenciador foi comercializado em 2015, o MinION, que trouxe consigo o conceito principal da praticidade, com o potencial em aplicações onde a portabilidade e espaço são limitados (SLATKO; GARDNER; & AUSUBEL, 2018). Além dele, a empresa lançou o GridION em 2018, PromethION em 2018, e outros equipamentos como Flongle e SmidhION.

Como dito anteriormente, essa tecnologia realiza o sequenciamento de moléculas únicas, logo, cada equipamento irá diferir do outro devido suas características (Tabela 5) (TECHNOLOGIES, 2021). Basicamente o uso de equipamentos maiores irá possibilitar o uso de mais *flow cells* funcionando ao mesmo tempo, consequentemente resultará em um sequenciamento em tempo real produzindo sequências longas com alto rendimento assim como a melhoria no tempo de sequenciamento. Segundo a própria Oxford Nanopore Technologies (2021) a maior sequência obtida até o momento foi de 4Mb. O fator limitador nessa tecnologia não está nela mesma e sim na quantidade/tamanho de moléculas a serem sequenciadas (DIJK et al., 2018).

Tabela 5 - Sequenciadores da Oxford Nanopore Technologies.

Empresa	Sequenciador	Nº de flow cells	Rendimento por flow cell (Gb)	Tempo de corrida	Nº de canais por flow cell	Custo do acesso ao sistema (Dólares)
	Flongle	1	1-2	1 min - 16h	126	\$ 1.460
Oxford Nanopore Technologies	MinION Mk1B	1	10 - 30 - 50	1 min - 48h	512	\$ 1.000
	MinION Mk1C	1	10 - 30 - 50	1 min - 48h	512	\$ 4.900
	GridION	5	10 - 30 - 50	1 min - 48h	512	\$ 49.995
	PromethION 24	24	100 - 200 - 300	1 min - 72h	2675	\$ 195.455

Fonte: <https://nanoporetech.com>.

O nanopore tem como fundamento o uso de proteínas transmembrana juntas a uma membrana lipídica que irão constituir poros, a partir da extremidade 5' da molécula (Figura 4). O adaptador junto a proteína motora desnatura o dsDNA. Após isso, o controle do ritmo de passagem do DNA pelo poro é comandado pela adição de proteínas motoras, uma DNA polimerase (phi29) que retarda o DNA devido a adição de nucleotídeos.

Essa passagem do DNA pelos poros acarreta em uma mudança de potencial isso porque existe em cada nanopore um eletrodo com conexão a um canal e um *chip* que medem a corrente elétrica que flui através do nanopore, nesse sentido, a medida que os nucleotídeos passam pelos poros produzem um sinal característico que interrompe a corrente elétrica gerando um sinal característico, esse sinal é decodificado junto a algoritmos que determinaram a sequência de DNA/RNA em tempo real (LIU et al., 2016); (SLATKO; GARDNER; & AUSUBEL, 2018).

Atualmente, o que se conhecia como *Next Generation Sequencing* (NGS), é denominado Sequenciamento de alto desempenho (HTS), termo que passou a ser utilizado amplamente na literatura.

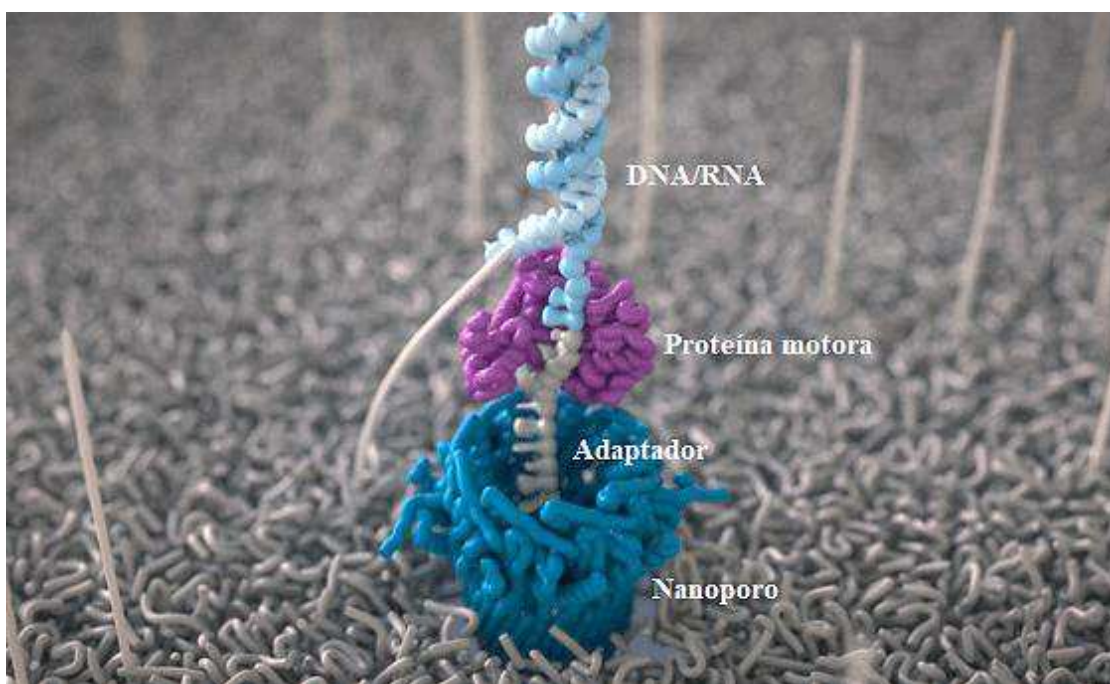


Figura 4 - Esquematização da tecnologia de sequenciamento nanopore da Oxford Nanopore Technologies. Fonte: Adaptado <https://nanoporetech.com/>.

Metodologia

Obtenção dos isolados

Os experimentos descritos nesse estudo foram desenvolvidos no laboratório de interação Planta-Praga III – LPP3 da Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF.

Estudos prévios de RCA-RFLP realizados por Boari e colaboradores (2019) foram realizados com DNA extraídos de folhas de coqueiros (*Cocos nucifera*) apresentando atrofia da coroa do coqueiro (ACC). Os resultados demonstraram uma associação de plantas apresentando sintomas da doença com componentes genômicos de um possível vírus multipartido, inicialmente associado à família *Nanoviridae*. (Boari e Ribeiro e dados não publicados). Em colaboração, as amostras de DNA dos isolados foram gentilmente cedidas pela Dra. Alessandra de Jesus Boari, pesquisadora da Embrapa Amazônia Oriental.

No Quadro 1 encontra-se a relação de amostras usadas nesse trabalho. Foram utilizadas amostras de DNA de coqueiros coletadas nos estados do Pará, Roraima, Bahia, Pernambuco e Paraíba, com sintomas (CS) de ACC ou sem sintomas aparentes (SS). Como controle, utilizamos uma amostra de coqueiro sadio e uma de palmeira areca (*Dyopsis lutescens*) procedente do Distrito Federal.

Quadro 1 - Relação de amostras de coqueiro. Número referente a cada amostra, local de coleta código da amostra, e presença de sintoma.

Amostras de Coqueiro			
Nº da amostra	Local de coleta	Código da amostra	Sintoma*
1	Moju/PA	1-P123 L113 P04	SS
2	Moju/PA	2-P123 L116 P05	SS
3	Moju/PA	3-P123 L114 P04	SS
4	Moju/PA	4-P114 L124 P06	CS
5	Moju/PA	5-P114 L123 P07	CS
6	Moju/PA	6-P114 L120 P01	CS
7	Moju/PA	7-P114 L121 P02	CS
8	Moju/PA	8-P114 L123 P02	CS
9	Moju/PA	9-P114 L123 P08	CS
10	Moju/PA	10-P114 L123 P05	CS
11	Sta Isabel/PA	21-H122 L13 P32	CS

12	Sta Isabel/PA	22- H122 L16 P33	CS
13	Sta Isabel/PA	23- H122 L19 P30	CS
14	Sta Isabel/PA	24- H122 L19 P29	CS
15	Sta Isabel/PA	25- H122 L23 P28	CS
16	Sta Isabel/PA	26- H122 L25 P31	CS
17	Sta Isabel/PA	27- H122 L25 P29	CS
18	Sta Isabel/PA	28- H122 L26 P29	CS
19	Sta Isabel/PA	29- H122 L29 P26	CS
20	Sta Isabel/PA	30- H122 L36 P25	CS
21	Capanema/PA	JRM PL1	CS
22	Capanema/PA	JRM PL2	CS
23	Capanema/PA	JRM PL3	CS
24	Capanema/PA	JRM PL4	CS
25	Capanema/PA	JRM PL7	SS
26	Una/BA	MT 1	CS
27	Una/BA	MT 2	CS
28	Una/BA	MT 3	CS
29	Una/BA	MT 4	SS
30	Una/BA	MT 5	SS
31	Una/BA	MT 6	SS
32	Una/BA	MT 7	CS
33	Una/BA	MT 8	CS
34	Una/BA	MT 9	CS
35	Una/BA	MT 10	CS
36	Una/BA	MT 11	CS
37	Una/BA	MT 12	SS
38	Una/BA	MT 13	SS
39	Una/BA	MT 14	SS
40	Canavieiras/BA	DUC 1	CS
41	Canavieiras/BA	DUC 2	CS
42	Canavieiras/BA	DUC 3	CS
43	Canavieiras/BA	DUC 4	SS
44	Canavieiras/BA	DUC 5	CS
45	Canavieiras/BA	DUC 6	CS
46	Canavieiras/BA	DUC 7	CS
47	Canavieiras/BA	DUC 8	CS
48	Canavieiras/BA	DUC 9	CS
49	Canavieiras/BA	DUC 10	CS

50	Canavieiras/BA	DUC 11	CS
51	Canavieiras/BA	DUC 12	CS
52	Canavieiras/BA	DUC 13	CS
53	Bom Jardim/PE	BJ1	CS
54	Bom Jardim/PE	BJ2	CS
55	Bom Jardim/PE	BJ3	CS
56	Bom Jardim/PE	BJ4	CS
57	Bom Jardim/PE	BJ5	CS
58	Bom Jardim/PE	BJ6	CS
59	Bom Jardim/PE	BJ7	CS
60	Bom Jardim/PE	BJ8	CS
61	Bom Jardim/PE	BJ9	CS
62	Bom Jardim/PE	BJ10	CS
63	Bom Jardim/PE	BJ11	CS
64	Bom Jardim/PE	BJ12	CS
65	Mamanguape/PB	NC1	CS
66	Mamanguape/PB	NC2	CS
67	Mamanguape/PB	NC3	SS
68	Mamanguape/PB	NC4	CS
69	Mamanguape/PB	NC6	CS
70	Mamanguape/PB	NC7	CS
71	Mamanguape/PB	NC8	CS
72	Mamanguape/PB	NC9	SS
73	Mamanguape/PB	NC10	SS
74	Mamanguape/PB	NC11	SS
75	Mamanguape/PB	NC12	SS
76	Lucena/PB	CV1	CS
77	Lucena/PB	CV2	CS
78	Lucena/PB	CV3	CS
79	Lucena/PB	CV4	SS
80	Lucena/PB	CV5	SS
81	Lucena/PB	CV6	CS
82	Lucena/PB	CV7	CS
83	Lucena/PB	CV8	CS
84	Lucena/PB	TB1	CS
85	Lucena/PB	TB2	CS
86	Lucena/PB	TB3	CS
87	Lucena/PB	TB4	CS
88	Boa Vista/RR	RRD1	CS
89	Boa Vista/RR	RRD2	CS
90	Boa Vista/RR	RRD3	CS
91	Boa Vista/RR	RRD4	CS
92	Boa Vista/RR	RRD5	CS
93	Boa Vista/RR	RRD6	CS

94	Belém/PA - Estufa – muda velha	Sadio	SS
95	Belém/PA - Estufa – muda nova	Sadio	SS

*SS (sem sintoma), CS (com sintoma)

Fonte: Autora.

Extração de DNA e Amplificação por círculo rolante (RCA)

A composição das soluções tampão utilizadas nesse trabalho estão detalhadas no Anexo I. As amostras previamente identificadas e separadas em tubos foram maceradas e o DNA foi extraído pelo método CTAB (GIBBS & MACKENZIE, 1997).

Foi adicionado 600 µL da solução de CTAB (2% CTAB; 1,4 M NaCl; 20 mM EDTA; 100 mM Tris-HCl, pH 6,8; 1% PVP e 0,1% β mercaptoetanol) em tubos de 1,5 ml e incubados 55°C por 20 minutos. Após isso foi adicionado 300 µL de clorofil 24:1, seguido de uma centrifugação (14.000 rpm) a 4°C por 5 minutos. A fase aquosa superior foi transferida para um novo tubo identificado. Foi adicionado 0,1 volume de acetato de amônio 7,5M e 1 volume de isopropanol (gelado).

Todos os tubos foram invertidos manualmente algumas vezes e então centrifugados por 30min a 14.000 RPM a 4°C. O sobrenadante foi descartado e o *pellet* foi lavado com 500 µL de etanol 70%. Secos a vácuo e ressuspensos em 50 µL de água Mili Q, incubado em banho seco a 37°C por 20 min e armazenado em freezer a -20°C.

Amostras de DNA total foram submetidas a *Rolling Circle Amplification* (RCA) (INOUE-NAGATA et al., 2004) através de uma reação contendo 0,1 µL de phi-29 DNA polimerase (BioLabs 10,000 U/mL), 1 µL do tampão (10X) da enzima phi-29 DNA polimerase (concentração final 1X), 1 µL de BSA (10X), 1 µL de dNTPs (10mM), 1 µL de primers randômicos para RCA (50 µM), 4,9 µL de água MilliQ e 1 µL de DNA total extraído, totalizando um volume final de 10 µL. A reação foi incubada por 48h a 37 °C. O produto da RCA obtido foi analisado em gel de agarose 1%, com migração a 100 volts.

Deteção do componente genômico Rep e purificação dos produtos de PCR

Duas amostras, 4 e 12 (quadro 1), ambas do estado do Pará foram selecionadas para a amplificação componente Rep completo para posterior clonagem, sequenciamento e caracterização molecular.

A reação de PCR se deu com 2,5 µL do tampão 10x *DreamTaq Buffer*, 0,5 µL de dNTP (10 mM), 0,25 de cada *primer* (Babu F e Babu R (Tabela 6)) (10 µM), 0,15 µL da *DreamTaq DNA polymerase* (Thermo Fisher Scientific 5U/ µL), 1 µL do produto da RCA e 20,35 µL de água milli-Q autoclavada, em um volume final de 25 µL. As condições da PCR foram 95°C por 3 minutos para uma desnaturação inicial, seguida por 30 ciclos de desnaturação a 95°C por 30 segundos, anelamento a 50°C por 30 segundos, extensão a 72°C por 1:30 minutos e um ciclo final de extensão a 72°C por 10 minutos.

Tabela 6 - Sequência dos *primers* utilizados para amplificação do componente Rep.

<i>Primers</i>	<i>Sequência</i>	<i>Referência</i>
Babu F	TCCTGAAGCTTACTCCCGTCACAA	KAUFFMANN et al., 2019;
Babu R	ACGATATCCTGAAGCTTACTTCCTGAAGCTTACT	BOARI et al., 2019.

Fonte: Autora.

Os produtos da PCR acrescidos de tampão de amostra foram aplicados em gel de agarose 1% em tampão TAE 1X (em anexo). As amostras foram submetidas a eletroforese com migração a 100V durante quarenta minutos e visualizados no fotodocumentador Gel Doc XR+ System (Bio-Rad). O fragmento de banda visualizada (~ 1.1kb) correspondente ao que se espera do componente Rep completo foi cortado com auxílio de lâmina purificado com o uso do kit Wizard® SV Gel and PCR Clean-Up System (Promega), seguindo as orientações do fabricante.

Clonagem e sequenciamento do componente Rep

Os fragmentos de DNA purificados foram clonados no vetor de clonagem pCR2.1_TOPO (3.9 kb) (*Invitrogen-Thermo Fisher Scientific*). A reação para inserção do DNA da Rep no plasmídeo se deu com a condição de 0,5 µL de *Salt Solution*, 0,5 µL do vetor, 1 µL de água milli-Q, e 1 µL de fragmento eluído, com tempo de incubação de trinta minutos em temperatura ambiente.

Os plasmídeos foram inseridos em células competentes de *Escherichia coli* cepa DH10B, pelo método de choque térmico. Tubos contendo 50 µL de célula competente

misturados com o plasmídeo foram incubados por 30 minutos no gelo. Posteriormente, os tubos foram colocados em banho seco a 42°C por 45 segundos e por fim, voltaram a incubação no gelo durante 2 minutos.

Posteriormente, foi adicionado 200 µL de LB (meio Luria-Bertani, em anexo) e a mistura foi incubada por 45 minutos a 37°C. O plaqueamento ocorreu em meio LB sólido contendo o antibiótico de seleção canamicina (concentração final 100µg/mL) e contendo X-Gal e IPTG, a placas foram incubadas a 37°C por 16 horas.

As colônias positivas selecionadas (coloração branca), provavelmente contendo o clone recombinante, foram repicadas e cultivadas em meio líquido contendo 3mL de LB acrescido de 3 µL de canamicina (concentração final 100µg/mL) por 16 horas sob agitação de 250 rpm e temperatura de 37°C.

Após o crescimento das culturas, foi realizada a preparação dos plasmídeos (miniprep) seguindo os seguintes passos descritos no Quadro 2:

Quadro 2 - Protocolo de miniprep (KOTCHONI et al., 2003).

1. 2mL de cultura foi transferida dos tubos falcon para eppendorfs de 1,5mL devidamente identificados;
2.Os mesmos foram centrifugados a 12000 rpm por 1 min;
3. O sobrenadante foi descartado e foi adicionado aos eppendorfs mais 0,5mL do meio de cultura;
4. Repetição do passo 2 e descarte do sobrenadante;
5. Foi adicionado aos eppendorfs 100µL da solução I (50 mM glucose, 10 mM EDTA e 25 mM Tris pH 8.0);
6. Utilizar agitador vortex para ressuspensão do pellet;
7. Adicionou-se 200µL da solução II (0.2 N NaOH e 1% SDS);
8. Os tubos foram gentilmente agitados a mão;
9. Acrescentar 150 µL da solução III (60 mL 5M KOAc, 5mL ácido acético 96% e 28,5mL de H2O) foi adicionada, o passo 8 foi repetido;
10. Incubar tubos no gelo por 5 minutos;
11.Centrifugar a 12000 rpm por 10 minutos;
12. 450uL do sobrenadante foram coletados e repassados para outros tubos eppendorfs de 1,5mL com suas respectivas identificações;
13. Foi adicionado o volume de 2x o sobrenadante de etanol absoluto e a mistura foi gentilmente agitada;
14. A mistura foi centrifugada a 12000 rpm por 15 minutos;
15. O sobrenadante foi descartado e foi adicionado 1mL de etanol 70% ao pellet;
16. Os tubos foram centrifugados por 2 minutos. O sobrenadante foi descartado e o pellet foi seco em uma capela de exaustão por 5 minutos;

- | |
|---|
| 17. O pellet foi ressuspensionado em 50µL de TE (0.1 M Tris pH 8.0 e 10 mM EDTA + RNase); |
| 18. Os tubos incubados a 65°C por 10 minutos. |

Após a realização da miniprep, os clones foram digeridos com a enzima de restrição EcoR I. Os clones recombinantes contendo o inserto foram confirmados e selecionados através da visualização em gel de agarose 1% do padrão de banda esperado.

Os clones selecionados foram cultivados novamente, a partir da cultura inicial, em 5mL de meio LB contendo 5 µL de Canamicina (concentração final 100µg/mL), os tubos foram incubados por 16h a 37°C sob agitação de 250 rpm.

Foi realizada uma extração do DNA plasmidial utilizando o kit Wizard® Plus SV Minipreps DNA Purification System (*Promega*), seguindo as recomendações do fabricante. Os clones (20 µL quantificados a 100ng/µL) foram sequenciados na empresa Macrogen (Coreia do Sul). Foram usados os primers universais T7 e M13 reverse para o sequenciamento nas duas orientações e obtenção da sequência completa do componente Rep.

Análise das sequências Rep obtidas dos Isolados 4 e 12

O genoma do componente Rep dos clones sequenciados foi montado utilizando o programa Geneious 11.0.5. A análise das sequências de nucleotídeos foi feita a partir de dados depositados no *National Center for Biotechnology Information* – NCBI, fazendo uso das ferramentas BLASTn e BLASTx (<http://www.ncbi.nlm.nih.gov/blast/>) além do software online *ORFinder* (<https://www.ncbi.nlm.nih.gov/orffinder/>) para identificação da ORF (*Open Reading frame*) que codifica a proteína do componente apontado como Rep.

Nesse trabalho, adotaremos o termo recomendado pela Dra. Alessandra Boari para o vírus em questão, “coconut crown atrophy virus” (CCAV). Os componentes do isolado 4 foram anteriormente clonados por enzimas de restrição (Boari e Ribeiro, dados não publicados). Foram identificados 7 possíveis componentes genômicos (6 por clonagens e a Rep por NGS) de um vírus multipartido nesse isolado que foram detectados apenas em amostras com ACC. Esses dados foram utilizados para comparação com os resultados obtidos no presente trabalho e caracterização dos isolados e dos componentes do CCAV e serão nomeados nos resultados por “coconut”.

Desenho de *primers* e amplificação de componentes genômicos em reação multiplex

As sequências de nucleotídeos dos seis componentes identificados do isolado 12 juntamente com as sequências dos componentes Rep do isolado 4 e 12 obtidas nesse trabalho foram alinhadas utilizando a ferramenta *multiple alignment* (MUSCLE alignment) no programa *Geneious* 11.0.5. Neste alinhamento foi possível a identificação de uma região de alta identidade entre todos os componentes e que representaria uma possível região comum maior (CR-M).

Com base na sequência da CR-M foram desenhados os pares de *primers* NanoCR1F/R e NanoCR1_2F/R (Tabela 7) com o intuito de amplificar todos os componentes genômicos do CCAV em uma reação de PCR inverso multiplex.

Baseado na frequência dos polimorfismos entre os componentes, foi preparado um mix dos primers na seguinte proporção $\frac{1}{4}$ do primer NanoCR1_2F (1,5 μ M): $\frac{3}{4}$ do primer NanoCR1F (8,5 μ M) e $\frac{1}{4}$ do primer NanoCR1_2R: $\frac{3}{4}$ (1,5 μ M) do primer NanoCR1R (8,5 μ M).

Inicialmente foi realizada uma PCR utilizando DNA total obtido das amostras de coqueiro do quadro 1. A reação foi composta por 1 μ L de DNA, 5 μ L de 10x SuperTaq Pol Reaction Buffer, 0,5 μ L dNTP, 2 μ L MgCl₂, 0,5 μ L do mix de primer forward (NanoCR1_2F+NanoCR1F) e 0,5 μ L do mix de primer reverse (NanoCR1_2R+NanoCR1R), 0,2 μ L da enzima SuperTaq Pol (Cellco), e 15,3 μ L de água milli-Q. Foi utilizado o programa utilizando as condições da PCR foram 95°C por 1 minutos, seguida por 35 ciclos de desnaturação a 95°C por 30 segundos, anelamento a 65°C por 30 segundos, extensão a 72°C por 1:30 minutos e um ciclo final de extensão a 72°C por 2 minutos. Posteriormente, a mesma reação foi feita com produtos da RCA das amostras de coqueiro.

Os produtos da PCR foram analisados em gel de agarose 1% em tampão TAE 1X. Os produtos da PCR junto ao de tampão de amostra foram aplicados no gel e submetidos a eletroforese, com migração a 100 volts, visualizados em fotodocumentador.

A partir do gel, os isolados foram escolhidos os isolados 4 e 12 como objetos de estudo para obtenção do componente genômico referente a Rep, estes foram clonados e sequenciados, assim como tínhamos acesso aos 6 componentes anteriormente clonados/sequenciados do isolado 12 (Boari e Ribeiro, dados não publicados). Devido a capacidade do mix de primers

NanoCR1 de acessar diversidade genética em amostras do Pará. Nesse sentido, no quadro 1 é possível observar a disponibilidade de três cidades do Pará onde foram obtidas amostras, o isolado 4 de Mojú/PA, isolado 12 de Sta. Isabel/PA e o isolado 22 de Capanema/PA, em vista disso, escolheremos como foco da dissertação a caracterização inicialmente limitada ao estado do Pará.

Uma nova PCR foi realizada com as amostras positivas, utilizando uma enzima de alta fidelidade, a Phusion™ High-Fidelity DNA Polymerase (Thermo Scientific™). A reação continha 5µL de 5x Phusion Buffer, 0,5µL dNTP (10 mM), 1,25µL do mix de *primer forward* (NanoCR1_2F+NanoCR1F) e 1,25µL do mix reverse NanoCR1_2R+NanoCR1R), 0,25µL da enzima Phusion, 0,5µL do produto da RCA e 16,25µL de água milli-Q, totalizando uma reação total de 25 µL. A reação se deu a partir da seguinte programação: 98°C por 30s, seguida por 25 ciclos de desnaturação a 98°C por 10 segundos, anelamento a 67°C por 30 segundos, extensão a 72°C por 35 segundos e um ciclo final de extensão a 72°C por 10 minutos.

A banda visualizada em gel de agarose correspondente ao fragmento amplificado de tamanho esperado para o componente genômico completo (1200 pb) foi cortada do gel com o auxílio de uma lâmina para realização da eluição. A purificação do DNA foi feita utilizando o kit Wizard® SV Gel and PCR Clean-Up System (Promega), seguindo as orientações do fabricante.

Tabela 7 - Sequência dos *primers* utilizados na PCR multiplex para obtenção dos componentes genômicos do CCAV.

<i>Primers</i>	Sequência
NanoCR1F	GGGATCGCGGAGTTCGACGG
NanoCR1R	TGAAAGATACGGGCGCTCGTCG
NanoCR1_2F	GGGGCCGCGGAGTCCGACGG
NanoCR1_2R	TGAAAGATACGGGGATCGTCG

Fonte: Autora.

Sequenciamento Nanopore e montagem dos componentes genômicos

As amostras amplificadas PCR utilizando os primers CR1-Mix (NanoCR1_2F+NanoCR1F e NanoCR1_2R+NanoCR1R) foram eluídas e posteriormente purificadas com o kit AMPure beads (Beckman Coulter, Estados Unidos), de acordo com as recomendações do fabricante. O DNA purificado foi quantificado com o kit Qubit dsDNA HS

(Life Technologies) no fluorímetro Qubit (Invitrogen). Cada amostra foi tratada com o NEBNext® Ultra™ II End Repair/dA-Tailing Module (NEB) para reparo das extremidades e ligadas aos adaptadores do ONT Native Barcoding Expansion kit (EXP-NBD104, Oxford Nanopore Technologies) com auxílio da enzima Blunt/TA Ligase Master Mix (New England BioLabs), de acordo com as recomendações do fabricante. Após inativação da enzima (65°C por 10 minutos), as diferentes amostras foram agrupadas em um único tubo e purificadas com o kit AMPure beads (Beckman Coulter, Estados Unidos). A biblioteca de sequenciamento foi gerada com o Ligation Sequencing Kit (SQK-LSK109 – Oxford Nanopore Technologies) e 20 ng do DNA resultante foi aplicado na *flow cell* (FLO-MIN106D - Oxford Nanopore Technologies) e o sequenciamento realizado no sequenciador MinION Mk1B (Oxford Nanopore Technologies) por cerca de 8 horas.

Os arquivos brutos FAST5 foram convertidos em FASTQ usando o *software* Guppy no modo de alta precisão (<https://nanoporetech.com/>). As sequências obtidas foram demultiplexadas para agrupar as amostras de acordo com os *barcodes* utilizados com o auxílio do software MinKNOW (Oxford Nanopore Technologies). As sequências relacionadas a um mesmo componente foram agrupadas com o programa Minimap2 (<https://github.com/lh3/minimap2>), os *contigs* montados *de novo* com o programa Flye (<https://github.com/fenderglass/Flye>) e melhorados com o programa Medaka v1.2.3 (Oxford Nanopore Technologies). Os *contigs* finais foram anotados com o BLAST X implementado no programa Geneious 11.0.5 (Biomatters, Ltd.).

Análise filogenética dos componentes Rep do CCAV

Para realização das análises filogenéticas, foram baixadas do *GenBank* as sequências nucleotídicas de componentes genômicos referentes aos componentes Rep de vírus pertencentes aos gêneros *Cocosatellite*, *Muscarsatellite*, *Clostunsatellite* e *Subclovsatellite* (família *Alphasatellitidae*), *Babuvirus* e *Nanovirus* (família *Nanoviridae*) (Tabela 8).

Tabela 8 - Vírus pertencentes aos gêneros *Cocosatellite*, *Muscarsatellite*, *Clostunsatellite* e *Subclovsatellite*, *Babuvirus*, e *Nanovirus* e números de acessos no *GenBank* dos componentes Rep utilizados nesse trabalho.

Gênero	Nº de acesso no GenBank	Vírus
<i>Cocosatellite</i>	M29963	Coconut foliar decay alphasatellite 1
	MF926426	Coconut foliar decay alphasatellite 2

	MF926429	Coconut foliar decay alphasatellite 4
	MF926430	Coconut foliar decay alphasatellite 5
<i>Muscarsatellite</i>	AF416471	Banana bunchy top alphasatellite 2
	HQ616080	Banana bunchy top alphasatellite 3
	KF435148	Cardamom bushy dwarf alphasatellite
<i>Clostunsatellite</i>	AB000922	Milk vetch dwarf alphasatellite 2
	KC979052	Pea necrotic yellow dwarf alphasatellite 2
	KX534409	Sophora yellow stunt alphasatellite 4
	KX534398	Sophora yellow stunt alphasatellite 5
	U16735	Subterranean clover stunt alphasatellite 2
<i>Subclovsatellite</i>	X80879	Faba bean necrotic yellows alphasatellite 1
	MF510471	Faba bean necrotic yellows virus associated alphasatellite 1
	KX534399	Sophora yellow stunt alphasatellite 1
<i>Babuvirus</i>	EF546813	Abaca bunchy top virus
	JX867550	Cardamom bushy dwarf virus
	S56276	Banana bunchy top virus
<i>Nanovirus</i>	KC978958	Black medic leaf roll virus
	GQ150778	Faba bean necrotic stunt virus
	AJ132180	Faba bean necrotic yellows virus
	AB027511	Milk vetch dwarf virus
	HE654123	Faba bean yellow leaf virus
	MF535450	Cow vetch latent virus
	GU553134	Pea necrotic yellow dwarf virus
	KC979054	Pea yellow stunt virus
	MK035728	Subterranean clover stunt virus
	MH048845	Sophora yellow stunt virus
	MK039132	Parsley severe stunt associated virus
A definir	Não publicado	Coconut crown atrophy virus 4
	Não publicado	Coconut crown atrophy virus 4 clone
	Não publicado	Coconut crown atrophy virus 12
	Não publicado	Coconut crown atrophy virus 12 clone
	Não publicado	Coconut crown atrophy virus 22
	Não publicado	Coconut crown atrophy virus 22 1

Para construção da árvore filogenética, as sequências codantes dos componentes Rep dos isolados de CCAV e de vírus pertencentes aos gêneros *Cocosatellite*, *Muscarsatellite*, *Clostunsatellite*, *Subclovsatellite*, *Babuvirus*, *Nanovirus* foram alinhadas com o algoritmo MAFFT (REF) (KATO et al., 2002). A inferência filogenética foi realizada pelo método de máxima verossimilhança e modelo de substituição GTR+F+I+G4, no software IQ-TREE (MINH et al., 2020). O suporte dos ramos foi estimado com o teste não Shimodaira-Hasegawa (SH-aLRT) (SHIMODAIRA; HASEGAWA, 1989).

Cálculo de identidades dos componentes dos diferentes isolados

A identidade entre as sequências nucleotídicas dos componentes genômicos comuns entre os isolados foi calculado usando programa SDT V1.2 (MUHIRE, VARSANI & MARTIN, 2014) a partir de um alinhamento obtido com o programa MUSCLE.

Identificação de *motifs* conservados da Rep

As sequências de aminoácidos (aa) da Rep de membros dos gêneros *Nanovirus* e *Babuvirus* (AZP73867.1, AZP73866.1, QCF45524.1, QPB75524.1, QPB75530.1, QGV56702.1, ADE87486.1, ATS17317.1, AVT42629.1, QLM02667.1, QIH29493.1, QBO55987.1, YP_010088040.1, AWB36049.1, ARI50280.1, AGG38935.1, YP_009508038.1, AGG38934.1, ARE67987.1, ALI57184.1, AKO71340.1) selecionadas do banco de dados *GenBank* foram agrupadas e alinhadas por gênero. O mesmo procedimento foi feito para os componentes genômicos Rep dos isolados de CCAV caracterizados nesse trabalho. O alinhamento das sequências foi feito no programa *Geneious 11.0.5* utilizando o programa MUSCLE. Para identificação dos *motifs* conservados RCR I, II, III (envolvidos na replicação do círculo rolante) e Walker A, B e C (domínios helicase) (KAZLAUSKAS; VARSANI; KRUPOVIC, 2018; STANTON, 2015) foi utilizado um logotipo da sequência (representação em consenso) fazendo uso do servidor online *WebLogo V3.4* (<https://weblogo.berkeley.edu/logo.cgi>) (CROOKS et al., 2004).

Resultados e discussão

Deteção, clonagem e análise molecular do componente genômico Rep dos isolados 4 e 12

A partir da descoberta de um DNA-Rep associado a família *Nanoviridae* por meio da técnica de sequenciamento de alto desempenho (HTS) em coqueiros apresentando atrofia da coroa do coqueiro no estado do Pará (BOARI et al., 2019; KAUFFMANN et al., 2019), noventa e quatro amostras de DNA de coqueiros com e sem sintomas da atrofia da coroa do coqueiro dos estados do Pará, Bahia, Roraima, Pernambuco e Paraíba foram compartilhados pela DRA. Alessandra Boari para realização desse estudo.

Inicialmente, entre as 94 amostras de coco disponíveis foram escolhidas duas do estado do Pará, isolados 4 (Moju/PA) e 12 (Sta. Isabel /PA) para a clonagem do componente Rep.

Fazendo uso dos *primers* BabuR e BabuF otimizados para detectar o DNA-Rep, foi realizada uma PCR dos produtos da RCA dos isolados e com isso foi possível obter *amplicons* de aproximadamente 1.1kb nas duas amostras (Figura 5). O tamanho dos *amplicons* obtidos está de acordo com o descrito na literatura para o DNA-R, componente codificador da Rep para nanovírus e babuvírus, que pode variar de 970 a 1.1 kb (VETTEN et al., 2019). A proteína Rep está envolvida na replicação de todos os componentes do genoma de vírus de ssDNA (TIMCHENKO et al., 1999; ZHAO et al., 2019).

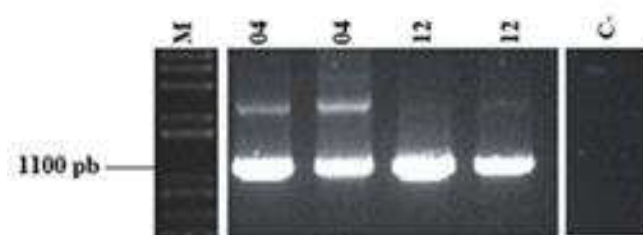


Figura 5 - Eletroforese em gel de agarose dos produtos da PCR com os primers BABUR e BABUF do RCA isolado das amostras de coco 4 e 12 do Pará. M- Marcador de peso molecular 1 Kb plus; C- Controle negativo.

Os produtos da PCR foram clonados no vetor pCR2.1_TOPO (3.9 pb). Na Figura 6 é possível observar os clones positivos obtidos dos isolados 4 e 12, de acordo com perfil de restrição com a enzima Eco RI. O inserto corresponde ao mesmo tamanho do *amplicon* de 1.100 pb utilizado na clonagem (Figura 6). Os clones foram sequenciados com os *primers* universais T7 e M13R.

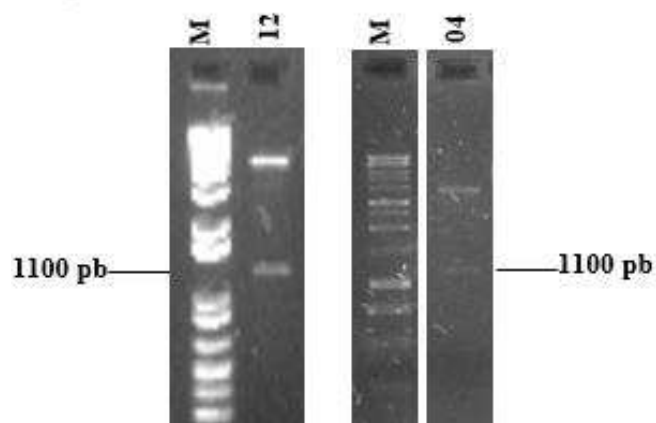


Figura 6 - Eletroforese em gel de agarose evidenciando padrão de restrição da digestão com a enzima Eco RI dos clones recombinantes dos isolados 04 e 12 em pCR2.1_TOPO. M- Marcador de peso molecular 1 Kb plus.

Os clones do componente Rep obtidos a partir do RCA dos isolados 4 e 12 de coco apresentando sintomas de atrofia da coroa do coqueiro, foram denominados coconut crown atrophy virus clone 4 (CCAV4CRep) e coconut crown atrophy vírus clone 12 (CCAV12CRep). Tanto o CCAV4C como o CCAV12C apresentam 1155 pb. Ao alinhar as sequências de nucleotídeos foi observado uma identidade de 98.095%. Isso fica evidente ao observar a Figura 7, onde os nucleotídeos em coloração vermelha apresentam 100% de identidade entre si, notavelmente esses abrangem praticamente toda extensão do alinhamento entre as sequências. Os nucleotídeos que estão em coloração branca apontam os locais onde ocorrem as diferenças entre as sequências, ao total foram 23 bases dos 1155 pb. Portanto, é possível supor que se trata do mesmo componente genômico de um mesmo vírus.

Utilizando o programa *ORFfinder* foi possível localizar as ORFs que caracterizam os clones em questão. Uma busca BLASTP da ORF do CCAV4CRep mostrou uma similaridade de 36.62% com “Master replication protein” de banana bunchy top virus (acesso Q86567.1] com uma cobertura de 87%, 34.75% de similaridade com “Master replication protein” de subterranean clover stunt virus (acesso Q9ICP7.1) com uma cobertura de 87% e 36.46% com a “Master replication protein” de faba bean necrotic yellows virus (acesso O39828.1) com 88% de cobertura.

Com relação a análise da ORF do CCAV12CRep, os resultados indicaram uma relação bem similar à do isolado 4 com identidade de 35.92% com a “Master replication protein” de banana bunchy top virus, acesso Q86567.1 com cobertura de 87%, também se relacionou com

a “Master replication protein” de subterranean clover stunt virus (acesso Q9ICP7.1) com identidade de 34.04% e cobertura de 87%, e com “Master replication protein” de faba bean necrotic yellows virus (acesso O39828.1)] apontou um perfil de identidade de 33.80% com cobertura de 88%.

A partir dessas análises é possível apontar que os clones CCAV4CRep e CCAV12CRep de fato são relacionados aos componentes genômicos Rep de membros da família *Nanoviridae* corroborando os resultados encontrados anteriormente por Boari e colaboradores (2019) a partir de *contig* obtido de sequenciamento HTS de amostras de coco apresentando atrofia da coroa do coqueiro.

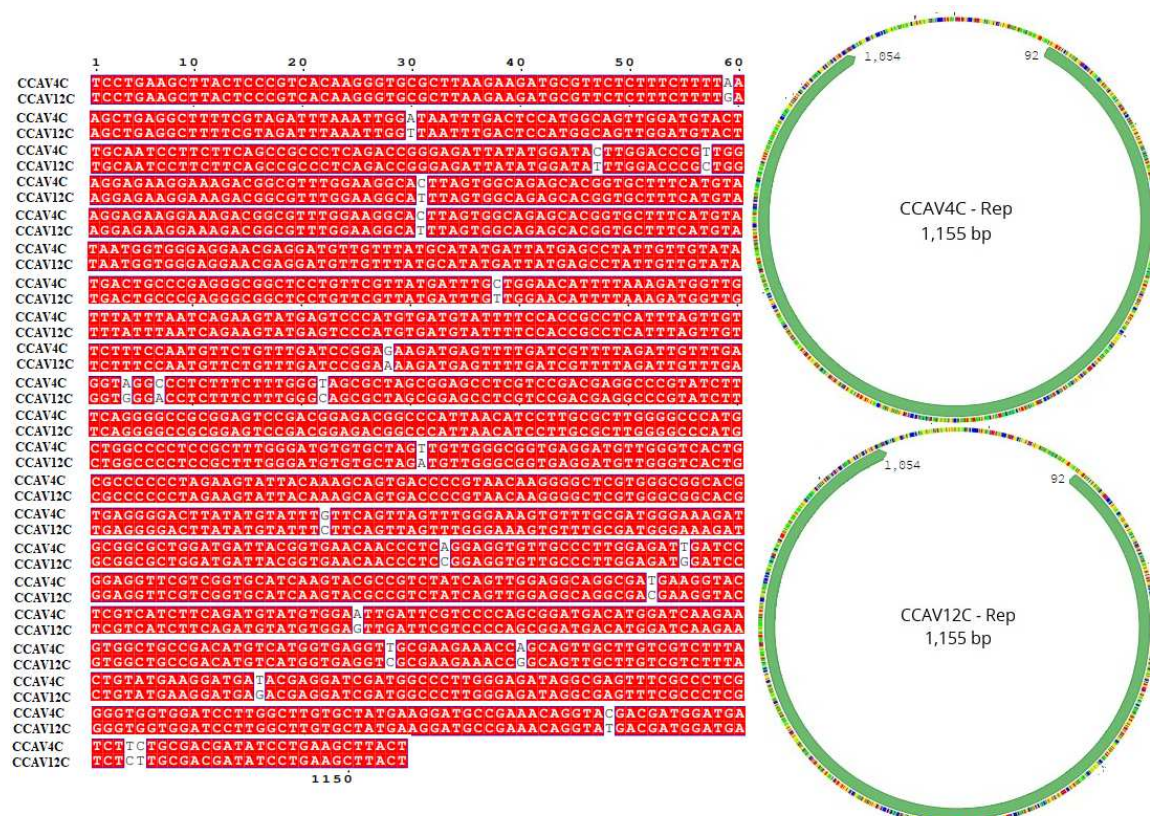


Figura 7 - Alinhamento da sequência de nucleotídeos dos clones virais 4 e 12. Em vermelho são evidenciados os nucleotídeos com 100% de identidade e ao lado representação dos componentes Rep, em verde encontram-se as ORF's determinadas pelo programa ORFfinder.

Desenho de primer para acesso à diversidade genética do coconut crown atrophy virus (CCAV)

A partir do sequenciamento, obtenção e confirmação que os clones dos isolados 4 e 12 de fato são identificados como uma Rep com associação a vírus multipartidos de ssDNA, foi possível iniciar o processo de caracterização molecular de três isolados de coconut crown

atrophy virus (CCAV). O primeiro passo se deu a partir da localização de características primordiais para um vírus “Cress DNA”. Esse termo é aplicado para vírus circulares de DNA de fita simples (ssDNA) que codificam uma proteína Rep associada a replicação (ZHAO et al., 2019).

Todos os vírus de ssDNA necessitam de uma proteína que inicie o processo de replicação, essa proteína é codificada por um componente genômico específico, como é o caso dos vírus da família *Nanoviridae* ou pelo DNA-A como ocorre na família *Geminiviridae* (GRONENBORN et al., 2018). Nos membros da família *Nanoviridae*, o DNA-R codifica a proteína Rep e compartilha sequências perto da origem de replicação com todos os DNAs componentes genômicos, cuja replicação depende da ação M-Rep (TIMCHENKO; KATUL; SANO, 2000).

Com essas informações, as sequências dos componentes Rep dos isolados 4 e 12 caracterizadas nesse trabalho foram alinhadas com as sequências de clones dos seis componentes genômicos do isolado 12, previamente clonados a partir de enzimas de restrição e sequenciados (Boari e Ribeiro, dados não publicados). Com isso foi possível localizar a região comum maior (CR-M) e fazer uso de uma parte dessa região (comum para todos os componentes genômicos) para o desenho de *primers* específicos.

Como diferentes componentes apresentaram polimorfismos na região escolhida, foram desenhados dois pares de primers (NanoCR1_F/NanoCR1R e NanoCR1_2F/NanoCR1_2R) (Figura 8, Tabela 7) representativos das respectivas variantes de sequência para que fosse possível à amplificação de todos os componentes do CCAV em uma reação de PCR multiplex (PCRm). A PCRm se baseia no uso dois ou mais pares de primers para detecção e amplificação de várias sequências de um determinado DNA alvo, em uma única reação (THOMPSON et al., 2003). O método tem como benefícios a redução de tempo, custo e esforço para diagnóstico de patógenos de humanos e animais sendo também amplamente utilizado para diagnose de patógenos em plantas (BERTOLINI et al., 2001; GALVEZ et al., 2020).

A PCRm é um método frequentemente usado quando se trata da identificação de diversos vírus infectando plantas, facilitando o diagnóstico prévio de doenças. PCRm já foi utilizado, por exemplo, para detecção simultânea do banana bract mosaic virus e sugarcane mosaic virus em abacá (*Musa textilis*) (GALVEZ et al., 2020). Em uma situação mais complexa, a PCRm foi usada para detecção de dois vírus e dois viróides, (grapevine fleck virus, grapevine

Pinot gris virus, grapevine yellow speckle viroid 1, e hop stunt viroid) em videiras (*Vitis sp.*) na República Tcheca (KOMÍNKOVÁ; KOMÍNEK, 2019). No Brasil, foi desenvolvido um protocolo de PCRM para detectar os dois vírus associados a "papaya sticky disease (PSD) ou “meleira do mamoeiro” (papaya meleira virus 1 e papaya meleira virus 2) doença que atinge severamente o mamoeiro, e proporcionou uma diagnose precoce da doença em uma única reação (MAURASTONI et al., 2020).

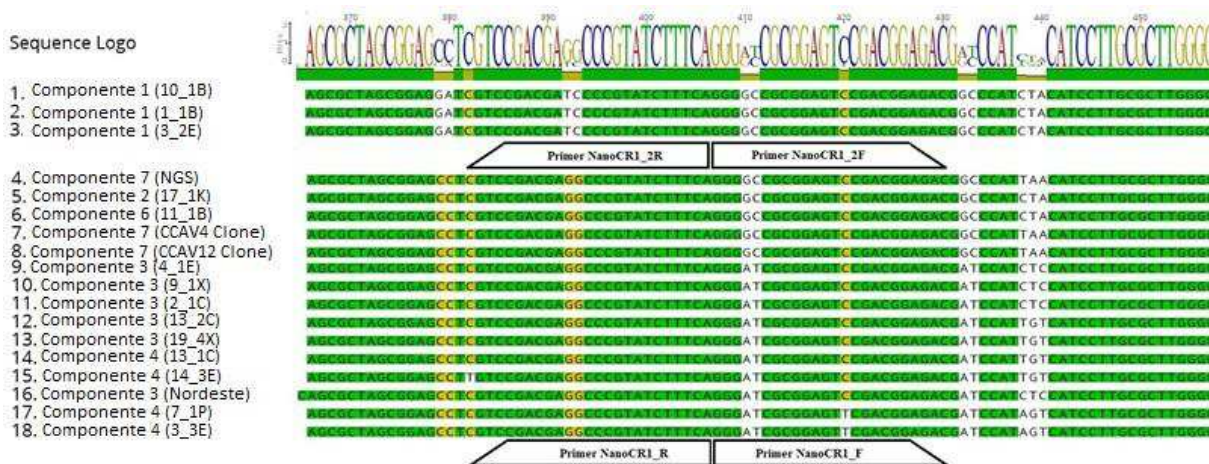


Figura 8 - Sequências dos componentes genômicos DNA-1, 2 ,3 ,4, 6 e 7 utilizadas no desenho dos *primers* para PCR multiplex. Fonte: Autora.

Diversidade genética do coconut crown atrophy virus (CCAV)

Inicialmente, reações de PCR utilizando o mix de primers NanoCR1 (NanoCR1_F, NanoCR1R, NanoCR1_2F e NanoCR1_2R) foram realizadas para triagem das amostras disponíveis (Quadro 1) e validação para o uso dos primers (Figura 9). Primeiramente, pela facilidade foi utilizado o DNA total extraído dessas amostras. Apenas quatro das 94 amostras apresentam o perfil de banda esperado de 1.100 pb, as amostras 06, 14, 15 e 17, todas coletadas no estado do Pará. Entretanto, as amostras 04 e 12 a partir das quais pudemos amplificar (com primers BabuF/BabuR), clonar e caracterizar o componente Rep não apresentaram fragmentos na reação de PCR com os novos *primers*.

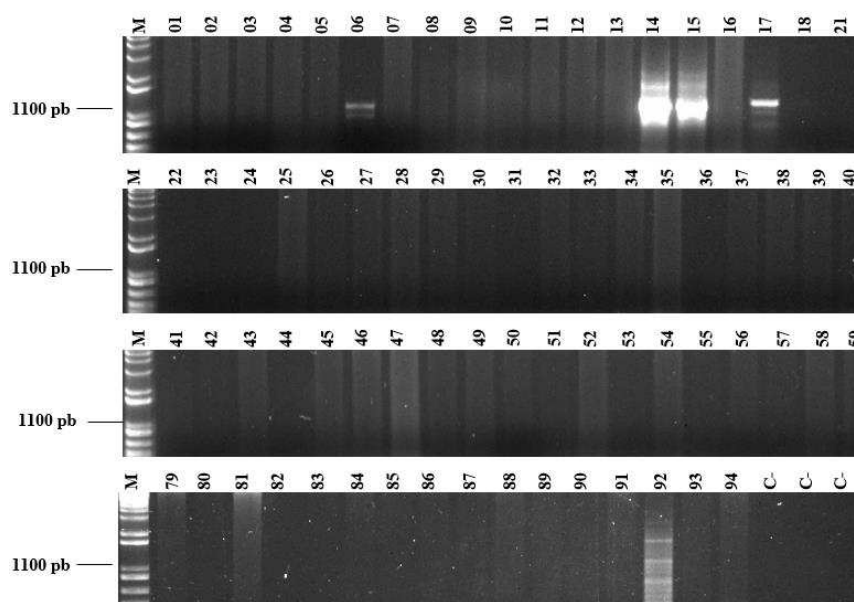


Figura 9 - Eletroforese em gel de agarose de fragmentos resultantes da PCR com os primers CR1-MIX com DNAs isolados das amostras de coqueiro Marcador de peso molecular 1 Kb plus; C- Controle negativo.

Esse resultado possivelmente se deve ao provável baixo número de cópias de ssDNA do CCAV nas amostras de DNA analisadas, que pode limitar a sensibilidade da técnica utilizada para detecção, podendo produzir resultados falsos negativos (TAKAHASHI et al., 2019). A presença de algum inibidor nas preparações de DNA também podem ter contribuído para o resultado negativo nas reações de PCR. Esses inibidores podem ter origem na amostra, ou durante o processamento/extração do DNA, levando a diminuição da sensibilidade da técnica ou resultados falso-negativos, ocasionando uma inibição parcial ou total da PCR (SCHRADER et al., 2012). Com isso, todas as mostras passaram por um enriquecimento de seu DNA viral por uma reação de RCA. A amplificação por círculo rolante (RCA) é um método bastante aplicado quando a carga de DNA viral é baixa, e resulta em uma síntese rápida de milhares de cópias de moléculas circulares de DNA, facilitando a amplificação e clonagem de gemomas virais (INOUE-NAGATA et al., 2004).

Os resultados da PCR feita com os produtos das RCAs das amostras de coqueiro contrastaram de maneira significativa quando comparada a PCR realizada com o DNA total. No geral, 20 amostras apresentaram o *amplicon* esperado de 1.100 pb (Figura 10). A maioria dos resultados positivos representa as amostras coletadas no Pará (04 a 09 de Mojú/PA, 12 a 17 de Santa Isabel/PA, 22 e 24 de Capanema/PA) mas amostras de outras procedências na Bahia (28 de Una/BA e 43 de Canavieira/BA), Pernambuco (55 de Bom Jardim/PE) e Paraíba (70 de Mamanguape/PB e 77 - cidade não identificada) também apresentaram resultado positivo. As

amostras apresentavam sintomas quando foram coletadas em 2019), com excessão de apenas uma amostra assintomática, a 43 oriunda do estado da Bahia.

Este experimento evidenciou que há uma diversidade entre os isolados de CCAV, pois foi possível observar que o uso dos *primers* NanoCR1 Mix foi eficiente no acesso aos isolados do CCAV em amostras do estado do Pará, tendo em vista que das 16 das 20 amostras positivas são provenientes desse estado. Como o restante das amostras também parece ser positivas para o CCAV em experimento realizado pelo grupo da Dr. Alessandra Boari, torna-se válido levantar a hipótese que a clonagem e sequenciamento dos componentes de um maior número de isolados, poderá gerar dados com os quais será possível o desenho de novos *primers* que poderão ser mais efetivos para o uso em PCRM para os isolados da região Nordeste.

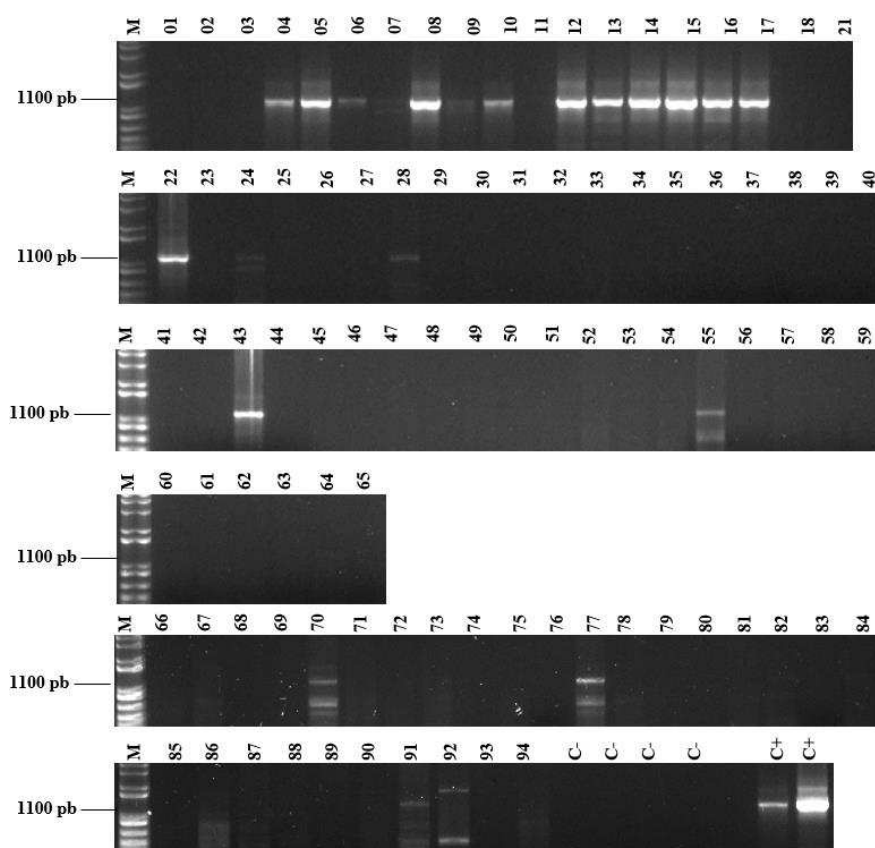


Figura 10 - Eletroforese em gel de agarose dos produtos da PCR com os primers CR1-Mix realizado com produtos de RCAs das amostras de coqueiro. M- Marcador de peso molecular 1 Kb plus; C- Controle negativo.

Os *amplicons* de uma segunda reação de PCRM das amostras que foram positivas foram purificados e visualizadas em gel para a constatação da pureza (Figura 11). Posteriormente, as amostras foram sequenciadas no sistema nanopore (MinION). Vale salientar que os resultados desta dissertação contêm os resultados obtidos para os isolados 04, 12 e 22 como foco para a

aplicação e validação da metodologia de PCRm e sequenciamento nanopore e caracterização molecular dos componentes desses isolados de CCAV.

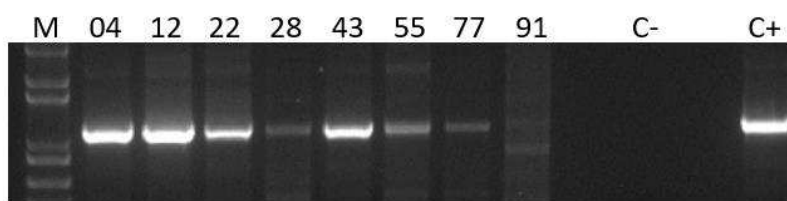


Figura 11 - Eletroforese em gel de agarose dos *amplicons* obtidos da PCRm com os primers CR1-Mix e purificados por eluição. M- Marcador de peso molecular 1 Kb plus; C- Controle negativo; C+ controle positivo.

Sequenciamento Nanopore aplicado a caracterização do CCAV

O uso da tecnologia de sequenciamento nanopore aplicado a identificação e monitoramento de vírus vem sendo amplamente utilizado nos últimos anos, especialmente nas epidemias e/ou pandemias como foi o caso do vírus Ebola, Zika e o novo coronavírus (JESUS et al., 2019; KONO; ARAKAWA, 2019; LIN; HUI; MAO, 2021; MUI et al., 2020).

Entretanto a tecnologia também se estende para outros fins. Sabe-se que para um controle mais efetivo de doenças em plantas, a detecção, identificação e conhecimento da diversidade do agente etiológico são fatores primordiais para a adoção das estratégias de controle. O sistema de sequenciamento nanopore pode contribuir em diversas dessas etapas. Por exemplo, o sequenciamento nanopore, a partir de amostras de DNA e RNA de plantas que passaram por um ensaio de inoculação com até três patógenos (vírus, fungos e bactérias) forneceu um diagnóstico inicial em tempo real (1-2 horas) após início do sequenciamento, possibilitando a classificação dos patógenos em nível de gênero e até mesmo de espécie (CHALUPOWICZ et al., 2019). Em outro estudo, o MinION foi aplicado ao “whole transcriptome amplification” WTA (amplificação do transcriptoma total) para o sequenciamento do metatranscriptoma de tecidos de plantas e insetos previamente infectados pelo plum pox virus, relatando a aplicação pioneira da tecnologia nanopore acoplada a WTA para diagnóstico em plantas e insetos vetores (BRONZATO BADIAL et al., 2018).

No presente estudo, no sequenciamento nanopore feito a partir dos produtos de PCR com uso dos primers NanoCR1 Mix (NanoCR1_F, NanoCR1R, NanoCR1_2F e NanoCR1_2R), foi possível identificar sete componentes genômicos em cada isolado de CCAV. Esses segmentos do isolado 12 já tinham sido previamente clonados e sequenciados pelo método de Sanger (Boari e Ribeiro, dados não publicados). Além dos 7 componentes do

CCAV de cada isolado (4, 12 e 22), um oitavo componente genômico foi identificado no isolado 22, totalizando ao final 22 componentes genômicos do coconut crown atrophy virus. Este resultado valida o uso dos primers NanoCR1 Mix na PCR multiplex para acessar a diversidade genética do CCAV, pelo menos nas amostras provenientes do estado do Pará.

Essa dissertação representa o primeiro relato da aplicação de PCR multiplex e sequenciamento nanopore como metodologia de identificação e caracterização do coconut crown atrophy virus em coqueiros (*Cocos nucifera* L.) apresentando atrofia da coroa do coqueiro.

Caracterização molecular dos genomas a partir do sequenciamento Nanopore

A análise das sequências dos componentes dos isolados do CCAV sequenciados pelo Nanopore confirmou a presença de duas regiões conservadas entre todos os componentes genômicos, CR-M (região comum maior) e CR-SL (região comum “stem-loop”). Cada componente do CCAV contém uma única caixa TATA, todos os segmentos abrigam sinais poli-A únicos (AATAAA). As características genômicas detalhadas de todos os segmentos estão resumidas na Tabela 9.

Tabela 9 - Caracterização dos componentes genômicos do coconut crown atrophy vírus. Fonte: Autora.

coconut crown atrophy virus (CCAV)	Componente genômico	ORF predita (Localização)	TATAbbox (Localização)	PolíA (Localização)	CR-M (Localização)	CR-SL (Localização)
Coconut Clone 12	1	255-710	204-207	799-804	838-909	1085-15
	2	148 - 720	159-162	770-775	856-926	1105-21
	3	158-592	105-108	607-612	757-827	1093-15
	4	142-615	*	940-945	816-875	1090-19
	5	263-715	224-227	738-743	864-934	1127-22
	6	182-472	*	482-487	553-623	707-26
CCAV12CRep	7	92-1054	53-56	843-838	980-1050	1127-19
CCAV4CRep	7	92-1054	53-56	843-838	980-1050	1134-26
CCAV4	1	255-710	204-209	799-804	839-909	1085-15
	2	141-713	152-155	763-768	849-919	1108-21
	3	158-595	105-108	607-612	724-824	1092-15
	4	142-618	*	1000-1005	815-885	1091-20
	5	263-715	224-227	738-743	864-934	1127-22
	6	136-834	*	844-849	915-985	1069-26
	7	92-1054	53-56	838-843	980-1050	1134-26
CCAV12	1	255-707	204-207	799-804	839-909	1085-15

	2	141-713	252-255	763-768	849-919	1110-21
	3	158-592	105-108	607-612	755-825	1093-15
	4	142-615	*	939-944	815-885	1089-19
	5	263-715	224-227	738-743	867-937	1130-22
	6	127-834	*	844-849	925-995	1079-26
	7	92-1054	53-56	838-843	980-1050	1134-26
CCAV22	1	255-707	204-207	799-804	839-909	1085-15
	2	141-713	252-255	763-768	849-919	1109-21
	3	158-592	105-108	607-612	757-827	1095-15
	4	142-615	*	939-944	815-885	1089-19
	5	263-715	224-227	738-743	866-936	1129-22
	6	134-841	*	851-856	932-1002	1079-26
	7	92-1054	53-56	838-843	980-1050	1118-19
	**7_1	93-1139	53-56	*	972-1044	1124-16

*Componentes genômicos que a TATAbox não foi localizada.

**Componente genômico 7_1 (Rep) a mais encontrado no isolado CCAV22, este será identificado como “CCAV227_1”

Em vírus multipartidos, como é o caso do BBTV, a CR-M e CR-SL são as duas regiões mais conservadas entre os segmentos do genoma (BURNS; HARDING; DALE, 1995). A CR-M (ou CR-II) possui um tamanho de geralmente 90 nucleotídeos (nt) podendo variar de isolado para isolado (BURNS; DALE; CENTRE, 1994). No CCAV foi encontrado uma região CR-M em torno de 70 nucleotídeos. As CR-Ms dos componentes genômicos dos isolados do CCAV foram alinhadas, utilizando o programa MAFFT e a partir do alinhamento foi obtido uma representação consensual das sequências fazendo uso do programa Weblogo v3.4 (Figura 12).

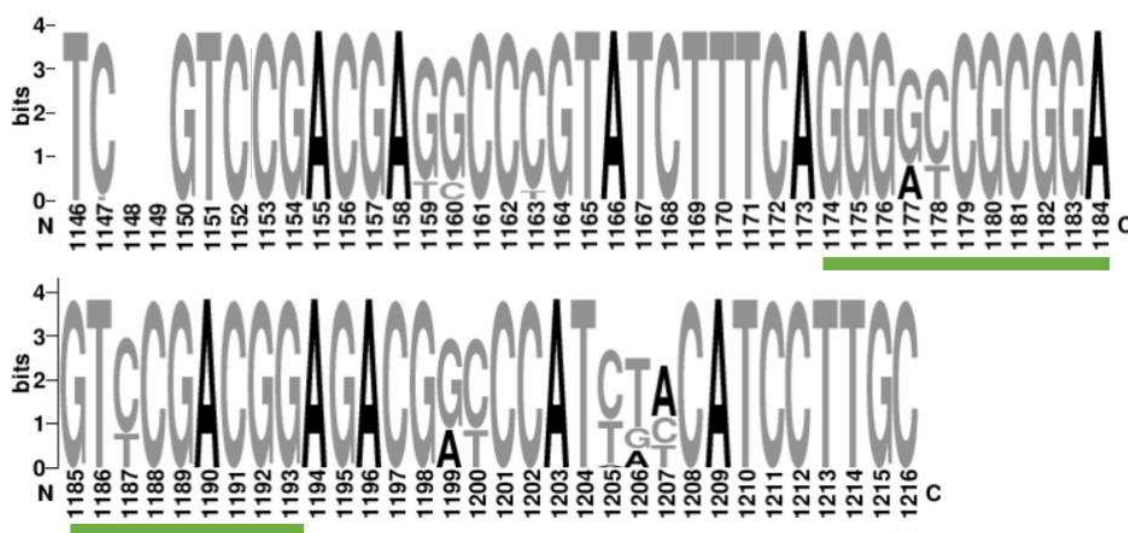


Figura 12 - Logotipo das sequências da CR-M alinhadas dos isolados de CCAV, em verde, a localização da região rica em CG na região comum maior. Fonte: Autora.

A partir disso foi possível identificar uma região da CR-M, em torno de 20 nucleotídeos, rica em GC (85%) (GGGGCCGCGGAGTCCGACGG) conservada em todos os componentes genômicos. Regiões ricas em GC pode exercer importantes funções. A sequência rica em GC da CR-M tem potencial de tornar a estrutura do DNA genômico de fita simples mais estável em BBTV (MUHIRE et al., 2014). Além disso, outras regiões ricas em GC ocorrem também em plasmídeos bacterianos com genoma de fita simples, com participação na fita complementar após a replicação por círculo rolante (RCR) (HAFNER; HARDING; DALE, 1997) Em geminivírus sabe-se da presença de regiões ricas em GC no elemento promotor de transcrição gene Rep1 (FENOLL et al., 1990).

Em todos os componentes genômicos, dentro da CR-SL foi encontrado o *stem loop*. Na figura 9, podemos observar uma representação dos *stem loops* para cada componente genômico dos isolados de CCAV. Os componentes comuns entre os isolados (DNA-1,2,3,4,5,6,7 e 7_1) (Figura 13).

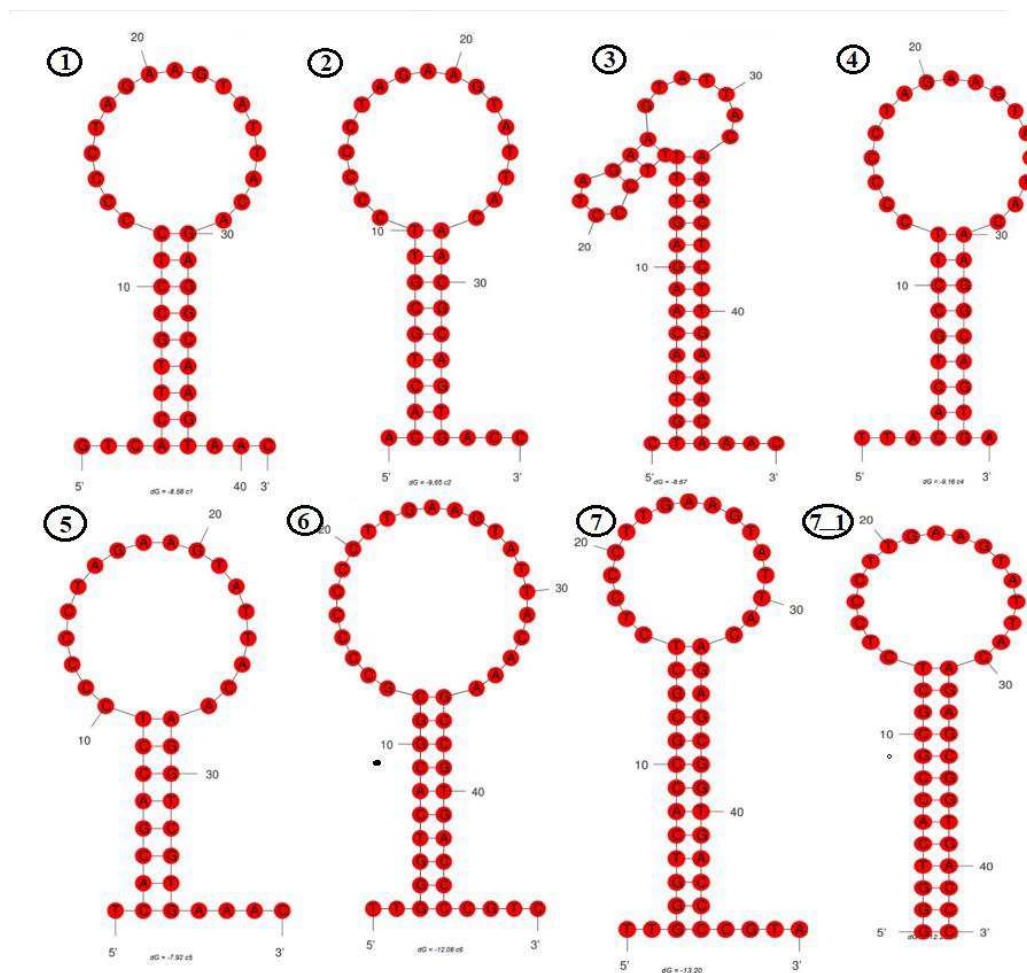


Figura 13 - Estrutura secundária do *hair-pin* dos componentes genômicos do CCAV, DNA-1, DNA-2, DNA-3, DNA-4, DNA5-, DNA6, DNA7 e DNA-7_1. Fonte: Autora.

A partir do alinhamento de todos os componentes dos isolados foi possível localizar o nonanucleotídeo 5'-AAGTATTAC-3' conservado em todos os DNAs do CCAV. Com isso, foi possível encontrar e prever a estrutura secundária da CR-SL. Os componentes apresentaram variação de tamanho nessa região. O DNA-1 tem 41 nt, DNA-2 40 nt, DNA-3 49 nt, DNA-4 39 nt, DNA-5 39 nt, DNA-6 49 nt, DNA-7 (Rep) 49 nt e o DNA-8 44 nt.

Com relação aos valores da energia livre de Gibbs (ΔG) das estruturas propostas dos oito componentes genômicos (DNA-1 a DNA-7_1), foram respectivamente: -4.32, -9.65, -8.67, -9.16, -7.92, -12.08, -13.20 e -6.68. A energia livre de Gibbs representa o potencial termodinâmico útil que é efetivamente utilizado na realização na reação ou trabalho. Os valores eletronegativos corroboram com o prognóstico de que as sequências naturalmente se moldariam nessas estruturas secundárias com menos gasto de energia.

O nonanucleotídeo conservado nos componentes do CCAV difere daqueles presentes em vírus dos gêneros *Babuvirus* (5'-TATTATTAC-3') e *Nanovirus* (5'-TAGTATTAC-3') (Thomas et al., 2021) e que representam a origem de replicação desses vírus.

Foram identificados sete componentes genômicos em cada isolado. Essa informação apresentada a partir do PCRm combinado com o sequenciamento nanopore era esperada tendo em vista a prévia clonagem de seis componentes do isolado 12 além do DNA-Rep dos isolados 12 e 4. A exceção ocorreu no isolado 22 que apresentou um componente genômico a mais, o DNA-7_1, que também codifica uma ORF com similaridade a Rep, sugerindo que o isolado 22 do CCAV possui dois componentes codificadores de Rep, mas existe a possibilidade da planta 22 estar infectada por dois vírus. Normalmente vírus membros da família *Nanoviridae* apresentam apenas uma Master-Rep. Durante a caracterização do *Sophora* yellow stunt-associated virus (SYSaV) foram identificados 4 componentes Rep em uma única planta (HEYDARNEJAD et al., 2017). Os autores sugerem que devido a presença de 4 componentes Rep divergentes entre si (identidade menor que 76,5%), a planta possa estar infectada por 4 vírus diferentes (HEYDARNEJAD et al., 2017).

Foi realizado um alinhamento múltiplo de cada um dos sete componentes genômicos do CCAV separadamente (Figura 14). As sequências denominadas COCONUT tem origem do isolado 12 previamente clonado por enzimas de restrição e sequenciado por Sanger (Boari e Ribeiro, dados não publicados), enquanto os componentes dos isolados 4, 12 e 22 foram sequenciados no nanopore a partir do uso da PCR multiplex com os primers CR1mix.



Figura 14 - Alinhamento múltiplo das sequências de nucleotídeos de cada componente genômico completo separadamente utilizando o programa MUSCLE. Em preto, as sequências e o consenso entre elas. Em verde, o perfil de identidade no decorrer do alinhamento. Fonte: Autora.

Na Figura 14 podemos observar que nos alinhamentos das sequências dos componentes 1, 2, 3, 4 e 5 dos isolados 4, 12 e 22 sequenciados no nanopore apresentam consenso e alta identidade ao longo de toda a sequência (linha em verde) quando comparados aos componentes genômicos do isolado 12 “COCONUT” sequenciados por Sanger. A identidade de nucleotídeos varia de 95-100% de identidade entre os componentes comuns.

Com relação ao componente genômico 6, é observado um “gap” de 662 nucleotídeos na sequência “COCONUT” do isolado 12 em relação aos isolados 4, 12 e 22. Esse gap ocorreu devido a presença de um segundo sítio da enzima de restrição BamHI e clonagem de um fragmento representando apenas uma parte do componente 6. Esse erro pode ser percebido e reparado no sequenciamento nanopore, onde a similaridade entre o componente 6 dos isolados desconsiderando esse gap fica em torno de 95-100% de identidade (Figura 15).

Com relação ao componente genômico 7, foram analisadas sete sequências, 4 e 12 (CCAV4CRep e CCAV12CRep) clonadas a partir de produto de PCR em PCR2.1 e sequenciadas por Sanger, descritas nesse trabalho, e os componentes 7 dos isolados 4, 12, 22.

Os componentes apresentaram uma porcentagem de 95-100% de identidade entre si. A exceção ocorreu no DNA-7_1 que apontou 81% de identidade com o componente 7 dos demais isolados e inclusive com o DNA-7 do próprio isolado 22.

Os resultados obtidos nesse estudo onde a sequência de 7 componentes foram obtidas de três isolados de CCAV e identidade entre os componentes comuns dos isolados é bastante alta (acima de 95%) sugere que as sequências genômicas se caracterizam como sendo provavelmente isolados distintos de um mesmo vírus. Ainda é válido salientar a eficácia da aplicação da metodologia de PCR multiplex seguido de sequenciamento nanopore aqui empregada para caracterização do CCAV em relação aos métodos de clonagem e sequenciamento tradicionais.

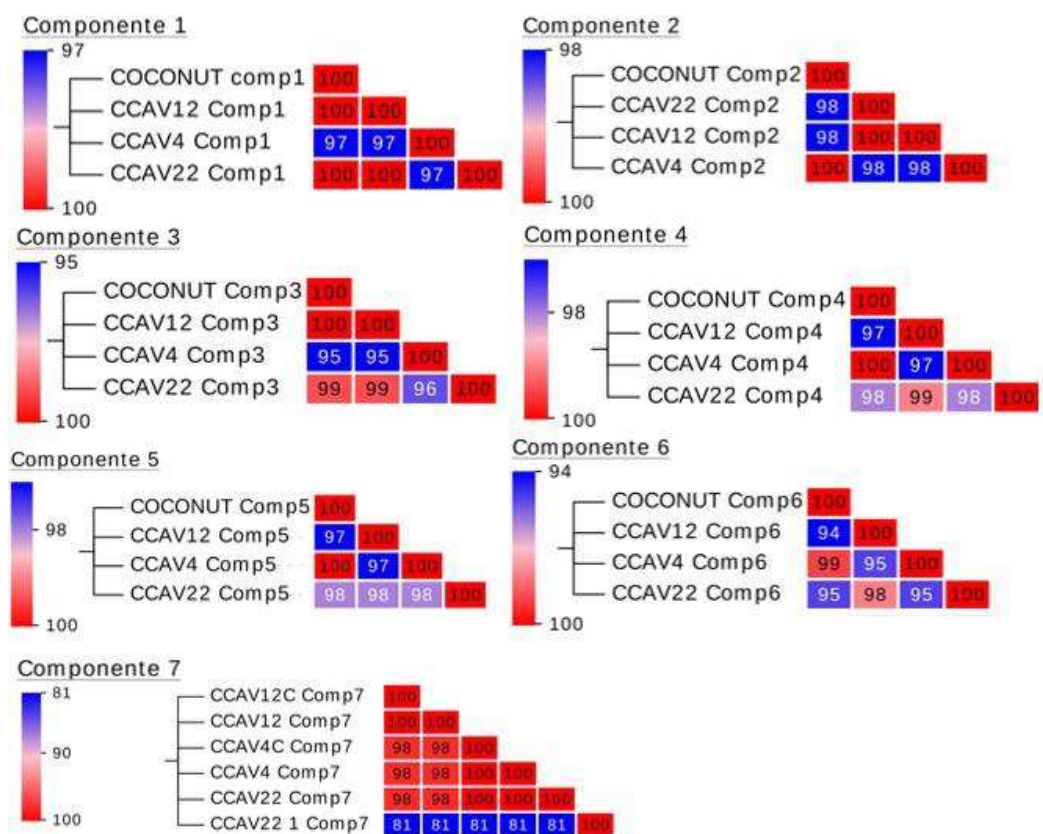


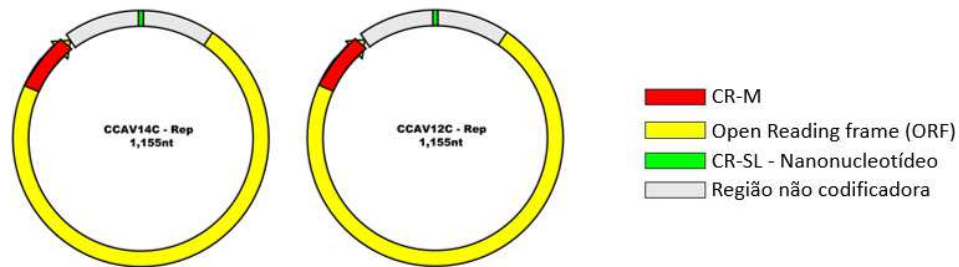
Figura 15 - Porcentagens de identidade da sequência de nucleotídeos dos DNAs genômicos do CCAV calculadas com o programa de demarcação de sequência (SDT). Fonte: Autora.

O coconut crown atrophy virus se caracteriza como um provável vírus de ssDNA circular que apresenta de 7-8 componentes genômicos com média de 1.1kb. Na Figura 16 observa-se a organização genômica dos componentes de cada isolado mostrando a posição das ORFs em sentido positivo, a CR-M, o nanonucleotídeo 5'-AAGTATTAC-3' presente na CR-

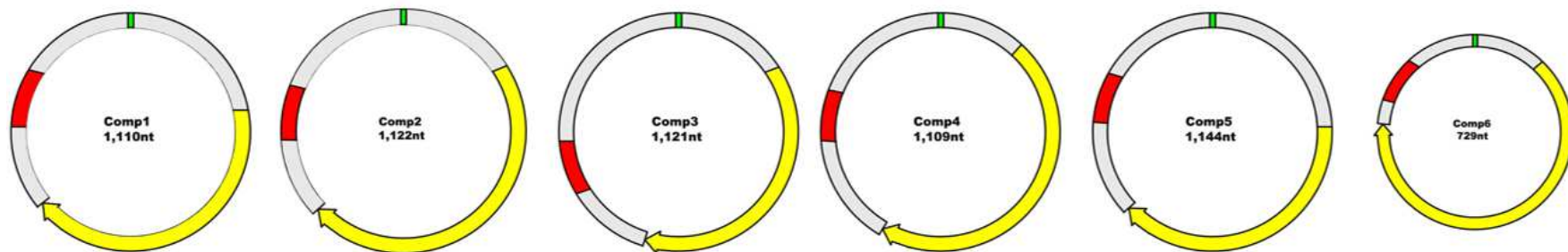
SL e a região não codificadora. Não foi possível até o momento identificar sequências conservadas na CR-SL em todos os componentes que representariam os *iterons*, que são sequências repetidas em sentido direto e/ou invertido que funcionam como sítios de reconhecimento e ligação da Rep ao DNA para iniciação da replicação (LONDOÑO; RIEGO-RUIZ; ARGÜELLO-ASTORGA, 2010). As ORFs de todos os componentes genômicos passaram por uma análise minuciosa que envolveu Blastn, Blastx, Blastp e blast-phi. Apenas o componente genômico 7 e 7_1 apresentaram identidade e similaridade com sequências virais relacionadas a Rep de nanovirids, como anteriormente visto no CCAV4CRep e CCAV12Rep, o componente 7 dos isolados 4,12 e 22 também caracterizados como um componente Rep. Não houve correspondência das proteínas codificadas pelas ORFs dos componentes 1-6 com nenhuma outra proteína presente no *GenBank*.

Estudos de caracterização biológica envolvendo a replicação dos componentes identificados nesse estudo pelo componente Rep são necessários para o estabelecimento definitivo da composição genômica do CCAV.

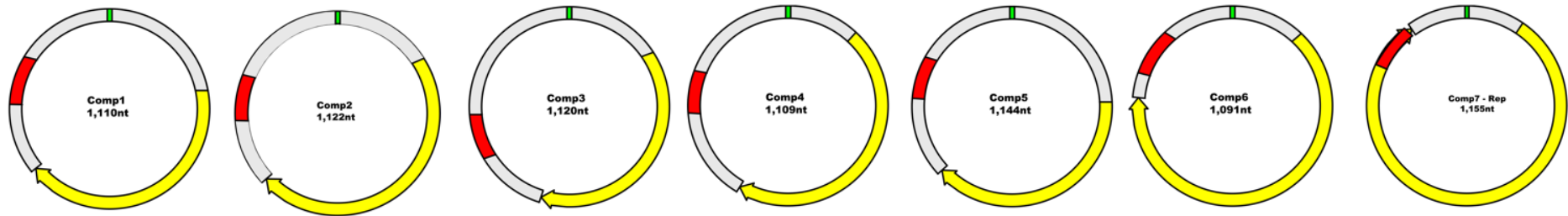
CCAV4Rep e CCAV12Rep – Clonagem por vetor – Sequenciamento Sanger



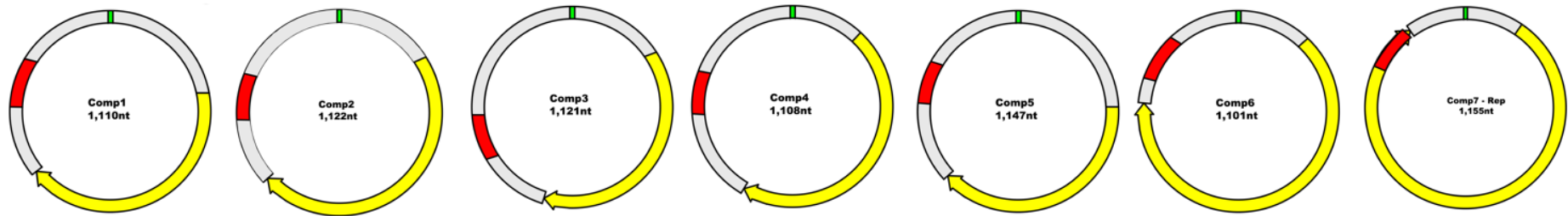
Coconut – Isolado 12 – Clonado por enzimas de restrição



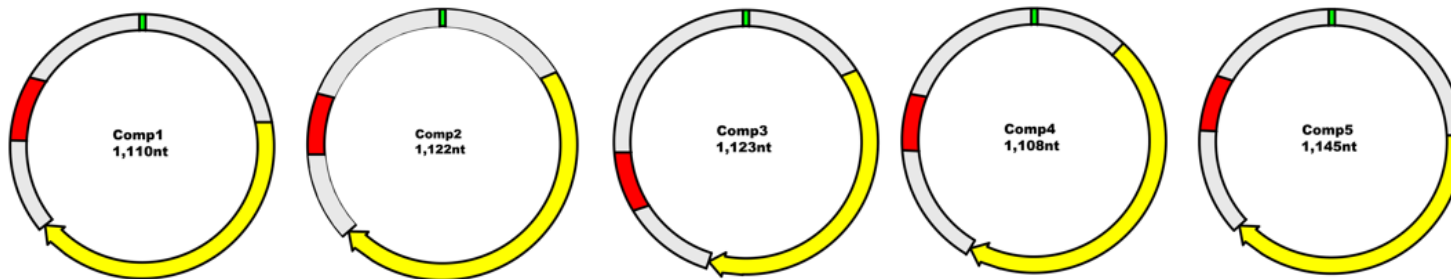
CCAV4



CCAV12



CCAV22



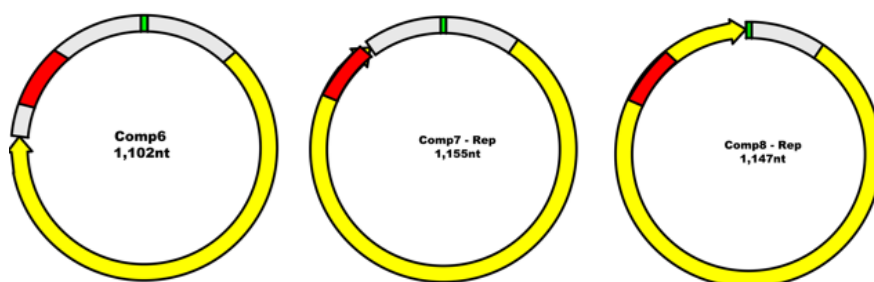


Figura 16 - Representação da caracterização molecular dos componentes genômicos do CCAV, em amarelo a representação da ORF dos componentes, em vermelho a região comum maior (CR-M) e em verde, a posição do *stem loop* (AAGTATTAC) Fonte: Autora.

Identidade e características conservadas em Reps do CCAV

Utilizando o programa MUSCLE, foi feito um alinhamento das proteínas codificadas pelas ORFs do componente genômico 7 dos isolados 4, 12 e 22. A similaridade obtida entre as sequências foi em torno de 95%. Uma das sequências deduzidas da proteína Rep foi utilizada em análise no software online Pfam, (SONNHAMMER; EDDY; DURBIN, 1997), com intuito de encontrar os domínios conservados relacionado a proteínas Rep. O software evidenciou uma região conservada (*Motif*) relacionada a helicase e outra região “Viral Rep”, relacionada a proteínas Rep envolvidas na replicação viral, e uma (Figura 17).

Motif in the sequence

Pfam ID:
RNA_helicase
Description:
PF00910, RNA helicase
Appearance:

Position	182..258
Alignment Query Database	IWILGPVGGEGKTAfGRHLVAE-----HGAFMYNGGRNEDVVYAYDYEPiVVYDCPRAAPVR--YDLEHFkdGCLFNQKYESH iwlyGps.geGKStlakeLakallkkklkklkksvysrnpeddfwdgYtgqevvvidDfgqvkddeepns.dlaelirlvsstpypln
i-Value	0.019

Sequence:
MGKMRRWITVNNPQEVLPLEIDPEVRRICKYAVYQLEAGDEGTRHLQMYVELIRPQRM
WIKKWLPTCHGEVAKKPAVACRLYCMKDDTRIDGPWEIGFRRPRGGGSLACAMKDAETGT
TMDLLRRYPEAYSRRHGALKKMRSLSFKAFAFRFKLDNLTPWQLDVLAILLQPPSORE
IIWILGPVGGEGKTAfGRHLVAEHGAFMYNGGRNEDVVYAYDYEPiVVYDCPRAAPVRyD
LLEHFkdGCLFNQKYESHVMYFPPPHLVVLSNVLFDPKMSDFRFLFEVGPLSLGSASG
ASSDEARIFQGPRSPTEAH

Motif in the sequence

Pfam ID:
Viral_Rep
Description:
PF02407, Putative viral replication protein
Appearance:

Position	5..88
Alignment Query Database	RRWITVNNPQEVLPLEIDPEVrrcIKYAVYQLEAGDEGTRHLQMYVELIRPQRMWIKKWLPT-CHGEVAKKPAVACRLYCMKD vswvFtlneyetellllvlqde....lkYlvyqdEradGtKhlgGfvefKkrtsLkqlkkllfpgrahLEiarGsqeearaYamKE
i-Value	5.1e-16

Sequence:
MGKMRRWITVNNPQEVLPLEIDPEVRRICKYAVYQLEAGDEGTRHLQMYVELIRPQRM
WIKKWLPTCHGEVAKKPAVACRLYCMKDDTRIDGPWEIGFRRPRGGGSLACAMKDAETGT
TMDLLRRYPEAYSRRHGALKKMRSLSFKAFAFRFKLDNLTPWQLDVLAILLQPPSORE
IIWILGPVGGEGKTAfGRHLVAEHGAFMYNGGRNEDVVYAYDYEPiVVYDCPRAAPVRyD
LLEHFkdGCLFNQKYESHVMYFPPPHLVVLSNVLFDPKMSDFRFLFEVGPLSLGSASG
ASSDEARIFQGPRSPTEAH

Figura 17 - Identificação a partir do software Pfam de motif relacionado a Rep nas sequências de proteína codificadas pelo componente 7 dos isolados 4, 12 e 22 do CCAV. Em vermelho, a região que o software apontou como motifs helicase e Viral_Rep.

Os vírus CRESS possuem seis *motifs* que são altamente conservados nas proteínas Rep, que estão envolvidos no processo de replicação (LONDOÑO; RIEGO-RUIZ; ARGÜELLO-ASTORGA, 2010). O RCR (replicação círculo rolante) I está envolvido no reconhecimento da origem de replicação, que ocorre pela associação a determinantes de especificidade (SPDs) (STANTON, 2015). Os SPDs são como sítios de reconhecimentos localizados próximo a sequências de *iterons* na CR-SL. O RCR II, domínio endonuclease HUH, está envolvido na coordenação de íons metálicos bivalentes, Mg 2+ ou Mn 2+, que são essenciais para a atividade da endonuclease na origem da replicação (LAUFS et al., 1995). O RCR III está envolvido na clivagem de dsDNA e subsequente ligação covalente do Rep através do resíduo catalítico de tirosina à extremidade 5' do produto clivado (KAZLAUSKAS et al., 2019; NASH et al., 2011)

Em geminivírus foi evidenciado que a região Walker A, B e C contém o *motif* helicase. Esse *motif* está envolvido no desenrolamento do DNA na direção 3'→5'. Pelas semelhanças em

funções e regiões conservadas, as regiões Walker A, B e C de Reps de vírus da família *Nanoviridae* também são considerados *motifs* de helicase (CHOUDHURY et al., 2006).

A partir do alinhamento de proteínas Rep de vírus pertencentes aos gêneros *Nanovirus* e *Babuvirus* (Família *Nanoviridae*) e as sequências das proteínas Rep dos componentes 7 e 7_1 do CCAV foi possível a identificação dos seis *motifs* Rep conservados. O alinhamento foi feito separadamente para cada grupo. A partir desse alinhamento foi criado um logotipo individual de cada alinhamento. A partir da localização dos *motifs* nos vírus da família *Nanoviridae* foi possível evidenciar os seis *motifs* conservados presentes nas Reps encontradas nos isolado 4, 12 e 22 do CCAV (Figura 18).



Figura 18 - Representação gráfica da variação em aminoácidos dos motifs RCR I, II e III e domínios helicase (Walker A, B e C) em representantes dos gêneros Babuvirus e Nanovirus e dos isolados de CCAV.

A Figura 19 ilustra a localização dos seis *motifs* conservados ao longo das sequências de aminoácidos das Reps dos componentes genômicos 7 e 7_1 do coconut crown atrophy virus.

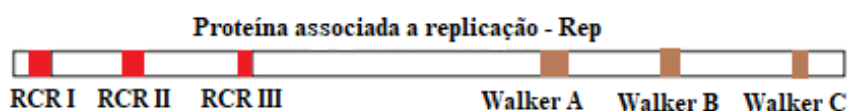


Figura 19 - Ilustração da localização dos seis *motifs* conservados na Rep de CCAV, com os três motivos RCR (replicação círculos rolante) e os motivos de helicase Waker A, B e C em marrom.

Análise filogenética das sequências Reps dos isolados do CCAV

O agrupamento filogenético revelou que as Reps dos diversos isolados de CCAV são relacionadas às Reps de membros da família *Nanoviridae* e *Alphasatellitidae* mas se agrupam em um clado separado dos membros das duas famílias (Figura 20).

Esse resultado indica, de forma ainda muito inicial, que o CCAV pode representar um novo táxon de vírus de genoma de ssDNA podendo representar um membro de um novo gênero da família *Nanoviridae* ou mesmo um membro de uma nova família de vírus multipartido de ssDNA de plantas. Porém devemos salientar que para a confirmação dessa hipótese, é necessário o estudo de mais isolados do CCAV e estudos de caracterização biológica que respondam perguntas indispensáveis para a confirmação da composição genômica e posição taxonômica desse vírus.

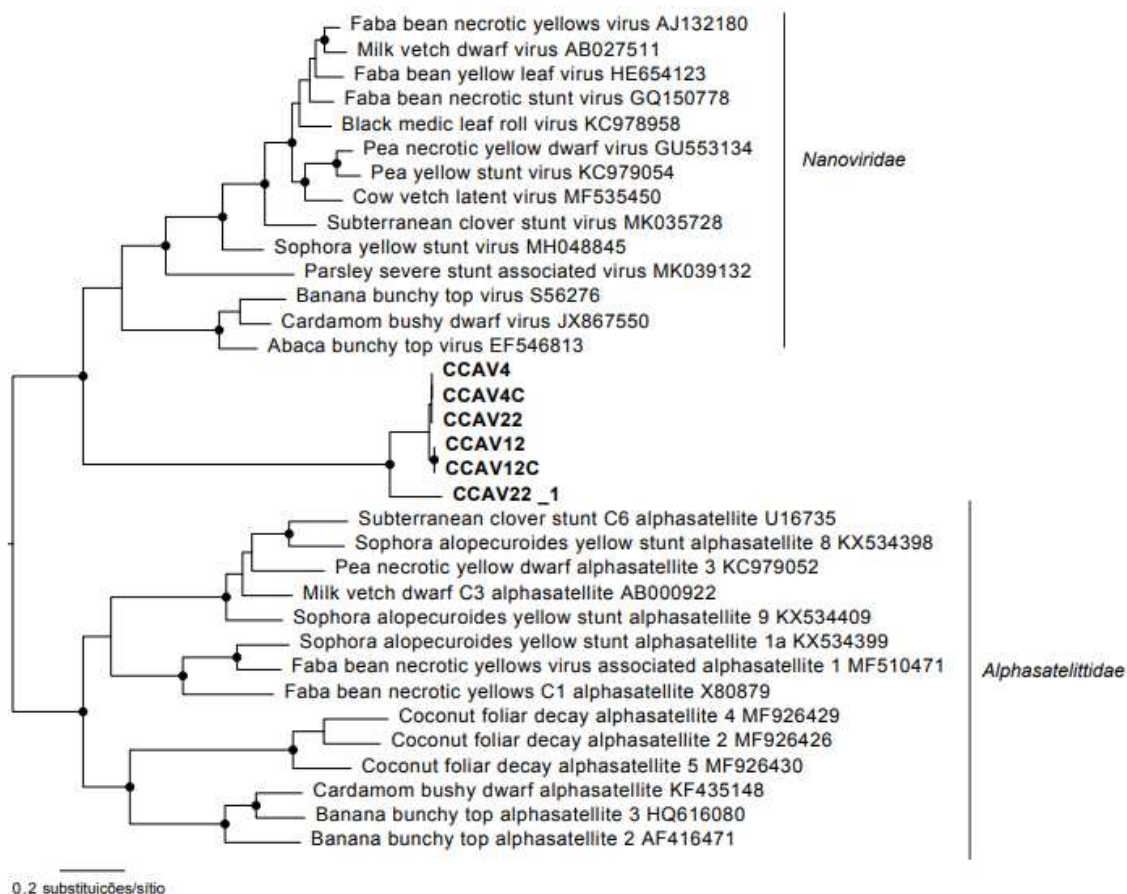


Figura 20 - Árvore filogenética de máxima verossimilhança construída a partir de um alinhamento de sequências codantes do componente Rep de vírus pertencentes aos gêneros *Cocosatellite*, *Muscarsatellite*, *Clostunsatellite* e *Subclovsatellite* (família *Alphasatellitidae*), *Babuvirus* e *Nanovirus*

(família *Nanoviridae*). A árvore foi enraizada por *midpoint*. Os círculos pretos mostram os ramos com suporte maior que 0,95.

Conclusões

- Os *primers* CR1-Mix aliados a PCR multiplex se mostraram eficientes no acesso a diversidade genética dos isolados de coconut crown atrophy virus (CCAV), especialmente dos estados do Pará;
- O sequenciamento nanopore mostrou-se como uma ferramenta eficaz, rápida e assertiva para caracterização molecular do CCAV frente a outras técnicas de clonagem e sequenciamento;
- Estudos filogenéticos sugerem que o CCAV é um membro de um novo táxon podendo representar um novo gênero da família *Nanoviridae* ou um membro de uma nova família de vírus de genoma de ssDNA de plantas

Perspectivas futuras

- Realizar a análise dos outros isolados obtidos a partir da PCR multiplex com uso dos *primers* CR-Mix e sequenciamento Nanopore;
- Realizar a caracterização molecular dos componentes encontrados em outros isolados para conseguir apontar se o CCAV de fato poderia se tratar de um novo gênero ou família de vírus de planta;
- Identificar se de fato o coconut crown atrophy virus é composto por sete componentes;
- Investigar a capacidade de autoreplicação da Rep e da Rep replicar os demais componentes dos isolados em tecidos foliares.
- Esse trabalho exhibe a caracterização inicial de isolados do CCAV, muitas perguntas ainda precisam ser respondidas com relação a esse vírus. Quando falamos em vírus multipartidos existem diversos desafios a serem vencidos, como por exemplo como ocorre a replicação; expressão e função das diferentes proteínas codificadas pelos componentes; localização nas células hospedeiras; sistema de transporte; identificação do vetor e modo de transmissão.

Anexo I

1- Tampão CTAB (Brometo de cetiltrimetilamônia)

Componente	Concentração	pH
CTAB	2% (p/v)	-
NaCl	1,4 M	-
Tris-HCl	100 mM	pH 8,0
β-mercaptoetanol	2% (v/v)	-
EDTA	20 mM	

Adaptado de (DOYLE; DOYLE., 1987).

3. Meio Luria-Bertani (LB)

Meio Luria-Bertani (LB)
Líquido
10 g de Triptona;
5 g de Extrato de levedura;
10 g de Cloreto de sódio;
Ajustar o volume para 1L com H ₂ O destilada q.s.p.;
Esterilizar por autoclavagem.
Sólido
Para 1L de meio LB líquido adicionar 16 g de ágar bacteriológico;
Esterilizar por autoclavagem.
Meio Luria-Bertani (LB) seletivo
1 µL Ampicilina (100 mg/mL)/ 1 mL de meio LB
1 µL Canamicina (100 mg/mL)/ 1 mL de meio LB
2 uL X-gal (20 mg/ mL)/ 1 mL de LB
0,2 uL de IPTG (0,8 M)/ 1 mL de LB

Fonte: (MILLER, 2010)

2- Solução TE

Componente	Concentração
Tris	10mM
EDTA	1Mm
dH ₂ O autoclavada	q.s.p. 100mL

3- Tampão TAE (Tris/ /EDTA) 50X

Componente	Concentração
------------	--------------

Tris	2M
EDTA	500mM
Ácido acético glacial	1M

4- Outras soluções

Solução	Concentração
Clorofil	Clorofórmio: Álcool Isoamílico 24:1
Tampão TAE (Tris-Acetato-EDTA) 1X	40 mM Tris-Acetato 1 mM EDTA

5- Antibiótico

Antibiótico	Dose seletiva (mg/L) <i>E. coli</i>	Solução estoque (mg/ml)	Solvente
Canamicina	100	100	Água

Referências

- ABRAHAM;VARRELMANN & VETTEN. Three distinct nanoviruses, one of which represents a new species, infect faba bean in Ethiopia. **Plant Disease**, v. 96, n. 7, p. 1045–1053, jul. 2012.
- ARAGAO, W. M.; ISBERNER, I. V.; CRUZ, E. M. DE O. **Água-de-coco**. [s.l.] Aracaju: Embrapa Tabuleiros Costeiros., 2001.
- ARONSON, M. N. et al. Clink, a Nanovirus-Encoded Protein, Binds both pRB and SKP1. **Jornal of Virology**, v. 74, n. 7, p. 2967–2972, 2000.
- BERTOLINI, E. et al. Single-step multiplex RT-PCR for simultaneous and colourimetric detection of six RNA viruses in olive trees. **Journal of Virological Methods**, v. 96, p. 33–41, 2001.
- BOARI, A. DE J. ; et al. Detecção do DNA-REP de Nanoviridae associado à Atrofia da Coroa do Coqueiro por meio de Sequenciamento de Nova Geração (Detection of the DNA-REP of Nanoviridae associated with the atrophy of the coconut crown by Next-Generation Sequencing). **Anais do IX Congresso Brasileiro de Fitopatologia**, p. 824, 2019.
- BRAINER, M. S. D. C. P.; XIMENES, L. F. X. Produção de Coco – Surgimento das Áreas Tradicionais do Nordeste. **Caderno setorial ETENE - Banco do Nordeste**, p. 1–15, 2020.
- BRONZATO BADIAL, A. et al. Nanopore Sequencing as a Surveillance Tool for Plant Pathogens in Plant and Insect Tissues. **Plant Disease**, v. 102, n. 8, p. 1648–1652, 2018.
- BURNS, R. M. H.; DALE, AND J. L.; CENTRE. Evidence that banana bunchy top virus has a multiple component genome. **Archives of Virology**, v. 137, p. 279–294, 1994.
- BURNS, T. M.; HARDING, R. M.; DALE, J. L. The genome organization of banana bunchy top virus : analysis of six ssDNA components. **Journal of General Virology**, v. 76, p. 1471–1482, 1995.
- CHALUPOWICZ, L. et al. Diagnosis of plant diseases using the Nanopore sequencing platform. **Plant Pathology**, v. 68, p. 229–238, 2019.
- CHOUDHURY, N. R. et al. The oligomeric Rep protein of Mungbean yellow mosaic India virus (MYMIV) is a likely replicative helicase. **Nucleic Acids Research**, v. 34, n. 21, p. 6362–6377, 2006.
- CHU, P. W. G. et al. Putative full-length clones of the genomic DNA segments of subterranean clover stunt virus and identification of the segment coding for the viral coat protein. **Virus Research**, v. 21, p. 161–171, 1993.
- CHU, P. W. G.; HELMS, K. Novel Virus-like Particles Containing Circular Single-Stranded with Subterranean Clover Stunt Disease DNAs Associated. **Virology**, v. 49, n. 1988, p. 38–49, 1989.

DIJK, E. L. VAN et al. The Third Revolution in Sequencing Technology. **Trends in Genetics**, v. 34, p. 666–681, 2018.

DINIZ, F. **Ácaro vermelho das palmeiras foi encontrado pela primeira vez no Brasil**. Disponível em: <http://www.infobibos.com/Artigos/2009_3/AcaroVermelho/index.htm>.

DOYLE, J. J.; DOYLE, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **Phytochemical Bulletin**, v. 19, p. 11–15, 1987.

FAJARDO, T. V. M. et al. High-throughput sequencing applied for the identification of viruses infecting grapevines in Brazil and genetic variability analysis. **Tropical Plant Pathology**, v. 42, p. 250–260, 2017.

FENOLL, C. et al. The intergenic region of maize streak virus contains a GC-rich element that activates rightward transcription and binds maize nuclear factors. **Plant Molecular Biology**, v. 15, n. 6, p. 865–877, 1990.

FERRAZ, L. G. Nova doença letal ao coqueiro no Brasil. **Tropical Plant Pathology**, v. 38, p. 583, 2012.

FERRAZ, L. G. B. et al. Nova doença ameaça coqueirais brasileiros. **Pesquisa Agropecuária Pernambucana**, v. 25, p. 1–4, 2020.

FERREIRA, J. M. S. ET AL. **A cultura do coqueiro no Brasil**. 2. ed. Brasília: Embrapa-SPI, 1998.

FERREIRA, S.; WARWICK, D.; SIQUEIRA, L. A CULTURA COQUEIRO NO BRASIL. **Embrapa**, v. 3, 2018.

FONTES, H. R.; FERREIRA, J. M. S. A cultura do coco. **Área de Informação da Sede-Col Criar Plantar ABC 500P/500R Saber (INFOTECA-E)**, 2006.

GAAFAR, Y. Z. A.; ZIEBELL, H. Aphid transmission of nanoviruses. **Archives of Insect Biochemistry and Physiology**, v. 104, n. 2, p. 1–13, 2020.

GAAFAR, Y.; ZIEBELL, H. Vicia faba, V. sativa and Lens culinaris as new hosts for Pea necrotic yellow dwarf virus in Germany and Austria. **New Disease Reports**, v. 34, p. 2044–0588, 2016.

GALLET, R. et al. Nanovirus-alphasatellite complex identified in Vicia cracca in the Rhône delta region of France. **Archives of Virology**, v. 163, n. 3, p. 695–700, 1 mar. 2018.

GALVEZ, L. C. et al. Multiplex reverse transcription-polymerase chain reaction for simultaneous detection of banana bract mosaic virus (BBrMV) and sugarcane mosaic virus (SCMV) in abaca Virology / Virologie Multiplex reverse transcription-polymerase chain reaction for sim. **Canadian Journal of Plant Pathology**, v. 92, p. 574–588, 2020.

GAMA, S.; ROBERTO, P.; FARIAS, S. Geostatistical modeling of “ lethal coconut palm crown atrophy ”, a new disease in the State of Pará. **Ciências Agrárias**, v. 42, p. 3603–3616, 2021.

- GIBBS & MACKENZIE, A. A primer pair for amplifying part of the genome of all potyvirids by RT-PCR. **Jornal of Virological Methods**, v. 63, p. 9–16, 1997.
- GOODWIN, S.; MCPHERSON, J. D.; MCCOMBIE, W. R. Coming of age : ten years of next-generation sequencing technologies. **Nature Publishing Group**, v. 17, n. 6, p. 333–351, 2016.
- GRIGORAS, I. et al. Reconstitution of Authentic Nanovirus from Multiple Cloned DNAs. **Journal of Virology**, v. 83, n. 20, p. 10778–10787, 15 out. 2009.
- GRIGORAS, I. et al. High Variability and Rapid Evolution of a Nanovirus. **Journal of Virology**, v. 84, n. 18, p. 9105–9117, 2010.
- GRIGORAS, I. et al. Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. **Journal of General Virology**, v. 95, p. 1178–1191, 2014a.
- GRIGORAS, I. et al. Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. **Journal of General Virology**, v. 95, p. 1178–1191, 2014b.
- GRIGORAS, I. et al. Nanovirus DNA-N encodes a protein mandatory for aphid transmission. **Virology**, v. 522, n. May, p. 281–291, 2018.
- GRONENBORN, B. et al. Analysis of DNAs associated with coconut foliar decay disease implicates a unique single-stranded DNA virus representing a new taxon. **Scientific reports**, v. 8, p. 1–14, 2018.
- GUŽVIĆ, M. The history of DNA sequencing. **Journal of Medical Biochemistry**, v. 32, n. 4, p. 301–312, 1 out. 2013.
- HAFNER, G. J.; HARDING, R. M.; DALE, J. L. A DNA primer associated with banana bunchy top virus. **Journal of General Virology**, v. 78, n. 2, p. 479–486, 1997.
- HEYDARNEJAD, J. et al. Identification of a nanovirus-alphasatellite complex in *Sophora alopecuroides*. **Virus Research**, v. 235, p. 24–32, 2017.
- HORSER, C. L.; HARDING, R. M.; DALE, J. L. Banana bunchy top nanovirus DNA-1 encodes the “master” replication initiation protein. **Journal of General Virology**, v. 82, n. 2, p. 459–464, 2001.
- IBGE. Área plantada ou destinada à colheita, área colhida, quantidade produzida, rendimento médio e valor da produção das lavouras temporárias e permanente. **Levantamento Sistemático da Produção Agrícola**, 2020.
- 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.
- INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES ICTV. Nanoviridae. **Virus Taxonomy: Nanoviridae**, 2021.
- JESUS, J. G. D. E. et al. Acute Vector-Borne Viral Infection : Zika and MinION Surveillance *. **Microbiology spectrum**, v. 7, p. 1–12, 2019.

KATOH, K. et al. MAFFT : a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic Acids Research**, v. 30, n. 14, p. 3059–3066, 2002.

KATUL, L. et al. Analysis of Six DNA Components of the Faba Bean Necrotic Yellows Virus Genome and Their Structural Affinity to Related Plant Virus Genomes 1. **Virology**, v. 233, p. 247–259, 1997.

KATUL, L. et al. Ten distinct circular ssDNA components , four of which encode putative replication-associated proteins , are associated with the faba bean necrotic yellows virus genome. **Jornal of General Virology**, v. 79, p. 3101–3109, 1998a.

KATUL, L. et al. Printed in Great Britain Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome. **Journal of General Virology**, v. 79, p. 3101–3109, 1998b.

KAUFFMANN, C. M. et al. Otimização da detecção do DNA-REP da família Nanoviridae em coqueiro com atrofia da coroa do coqueiro. **Anais do IX Congresso Brasileiro de Fitopatologia**, p. 788, 2019.

KAZLAUSKAS, D. et al. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. **Nature Communications**, v. 10, n. 1, p. 1–12, 2019.

KAZLAUSKAS, D.; VARSANI, A.; KRUPOVIC, M. Pervasive Chimerism in the Replication-Associated Proteins of Uncultured Single-Stranded DNA Viruses. **Viruses**, v. 10, p. 187–198, 2018.

KCHOUK, M.; GIBRAT, J. F.; ELLOUMI, M. Generations of Sequencing Technologies: From First to Next Generation. **Biology and Medicine**, v. 9, p. 1–8, 2017.

KOMÍNKOVÁ, M.; KOMÍNEK, P. Development and validation of RT-PCR multiplex detection of grapevine viruses and viroids in the Czech Republic. **Journal of Plant Pathology**, v. 102, p. 511–515, 2019.

KONO, N.; ARAKAWA, K. Nanopore sequencing : Review of potential applications in functional genomics. **Development, Growth & Differentiation**, v. 61, p. 316–326, 2019.

KOTCHONI, S. O. et al. A home made kit for plasmid DNA mini-preparation. **African Journal of Biotechnology**, v. 2, n. April, p. 88–90, 2003.

LAGEIX, S. et al. The Nanovirus-Encoded Clink Protein Affects Plant Cell Cycle Regulation through Interaction with the Retinoblastoma-Related Protein. **Journal of Virology**, v. 81, n. 8, p. 4177–4185, 2007.

LAUFS, J. et al. In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. **Biochemistry**, v. 92, n. April, p. 3879–3883, 1995.

LIMA, S. A. J. Avaliação físico-química, microbiológica e sensorial da água de coco anão verde comercializadas pelas indústrias do Sertão da Paraíba e do Ceará. 2013.

LIN, B.; HUI, J.; MAO, H. Nanopore Technology and Its Applications in Gene Sequencing. **Biosensors**, v. 11, p. 1–17, 2021.

LIU, Z. et al. Solid-State Nanopore-Based DNA Sequencing Technology. **Jornal of Nanomaterials**, v. 2016, p. 1–13, 2016.

LONDOÑO, A.; RIEGO-RUIZ, L.; ARGÜELLO-ASTORGA, G. R. DNA-binding specificity determinants of replication proteins encoded by eukaryotic ssDNA viruses are adjacent to widely separated RCR conserved motifs. **Archives of Virology**, v. 155, n. 7, p. 1033–1046, 2010.

MACHADO, L. J. et al. Utilização da biomassa do coco verde (cocos nucifera L.) Para obtenção de subprodutos. **Brazilian Journal of Development**, v. 6, n. 1, p. 3808–3826, 2020.

MAGALHÃES, NEY, T. et al. CONJUNTURA DE MERCADO DO COCO DA BAÍA (Cocos nucifera L.) NA REGIÃO AMAZÔNICA, COM ÊNFASE NO ESTADO DO PARÁ. Apresentação: Comunicação Oral. **II Congresso Internacional das Ciências Agrárias Cointer**, 2017.

MAKKOUK, K. M.; KUMARI, S. G. Epidemiology and integrated management of persistently transmitted aphid-borne viruses of legume and cereal crops in West Asia and North Africa. **Epidemiology and Integrated**, v. 141, p. 209–218, 2009.

MANDAL, B. et al. Nine novel DNA components associated with the foorkey disease of large cardamom : Evidence of a distinct babuvirus species in Nanoviridae. **Virus Research**, v. 178, n. 2, p. 297–305, 2013.

MAPA. **Estatísticas de Comércio Exterior do Agronegócio Brasileiro**. Disponível em: <<https://indicadores.agricultura.gov.br/agrostat/index.htm>>.

MARGULIES, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. **Nature**, v. 437, n. September, p. 376–381, 2005.

MAURASTONI, M. et al. A multiplex RT - PCR method to detect papaya meleira virus complex in adult pre - flowering plants. **Archives of Virology**, v. 111, p. 85–93, 2020.

MAXAM, A. M.; GILBERT, W. A new method for sequencing DNA. **Biochemistry**, v. 74, n. 2, p. 560–564, 1977.

MENGEL, KONRAD; KIRKBY, E. Principles of plant nutrition. **Springer Science & Business Media**, 2012.

MILLER, F. P. Molecular genetics. **Alphascript Publishing**, 2010.

MINH, B. Q. et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. **Molecular Biology and Evolution**, v. 37, n. 5, p. 1530–1534, 2020.

MOROZOV & SOLOVYEV, G. Small hydrophobic viral proteins involved in intercellular movement of diverse plant virus genomes. **AIMS Microbiology**, v. 6, n. 3, p. 305–329, 2020.

MUHIRE, B. M. et al. Evidence of Pervasive Biologically Functional Secondary Structures

within the Genomes of Eukaryotic Single-Stranded DNA Viruses. **Journal of Virology**, v. 88, n. 4, p. 1972–1989, 2014.

MUI, W. et al. Identification of nsp1 gene as the target of SARS - CoV - 2 real - time RT - PCR using nanopore whole - genome sequencing. **Journal of Medical Virology**, v. 92, p. 2725–2734, 2020.

NASH, T. E. et al. Functional Analysis of a Novel Motif Conserved across Geminivirus Rep Proteins. **Journal of Virology**, v. 85, n. 3, p. 1182–1192, 2011.

ORTIZ, V. et al. Incidence and transmission of Faba bean necrotic yellows virus (FBNYV) in Spain. **Span. J. Agric. Res**, v. 4, n. September, p. 255–260, 2006.

PRINGLE, C. R. Virus taxonomy–San Diego. **Archives of Virology**, v. 143, p. 1449–1459, 1998.

SANGER, F.; NICKLENX, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **National Academy of Sciences**, v. 74, n. 12, p. 5463–5467, 1977.

SANO, Y. et al. Sequences of ten circular ssDNA components associated with the milk vetch dwarf virus genome. **Journal of General Virology**, v. 79, p. 3111–3118, 1998.

SAUNDERS, K.; STANLEY, J. A Nanovirus-like DNA Component Associated with Yellow Vein Disease of *Ageratum conyzoides*: Evidence for Interfamilial Recombination between Plant DNA Viruses. **Virology**, v. 152, p. 142–152, 1999.

SCHRADER, C. et al. PCR inhibitors - occurrence, properties and removal. **Journal of Applied Microbiology**, v. 113, n. 5, p. 1014–1026, 2012.

SELVARAJAN, R.; BALASUBRAMANIAN, V. Host–virus interactions in banana-infecting viruses. **Plant Virus-Host Interaction: Academic Press**, p. 57–78, 2014.

SELVARAJAN, S. V. R. Recent advances in understanding the replication initiator protein of the ssDNA plant viruses of the family Nanoviridae. **Virus Disease**, v. 30, n. 1, p. 22–31, 2019.

SHARMAN, M. et al. Abacá bunchy top virus, a new member of the genus Babuvirus (family Nanoviridae). **Archives of Virology**, v. 153, p. 135–147, 2008.

SHIMODAIRA, H.; HASEGAWA, M. Letter to the Editor Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. **Molecular Biology and Evolution**, v. 16, p. 1114–1116, 1989.

SICARD, A. et al. A multicellular way of life for a multipartite virus. **eLIFE**, v. 8, p. 1–14, 2019.

SIQUEIRA, L. A.; ARAGÃO, W.; & TUPINAMBÁ, E. A introdução do coqueiro no Brasil, importância histórica e agrônômica. **Aracaju: Embrapa Tabuleiros Costeiros**, p. 24, 2002.

SLATKO, B. E.; GARDNER, A. F.; & AUSUBEL, F. M. Overview of Next- Generation Sequencing Technologies. **Current Protocols in Molecular Biology**, v. 122, p. 1–11, 2018.

SONNHAMMER, E. L. L.; EDDY, S. R.; DURBIN, R. Pfam: A comprehensive database of protein domain families based on seed alignments. **Proteins: Structure, Function, and Bioinformatic**, v. 28, p. 405–420, 1997.

STAINTON, D. B. **Towards understanding the evolution of Banana bunchy top virus and the detection of associated badnaviruses**. [s.l.] University of Canterbury., 2015.

TAKAHASHI, H. et al. Virus Latency and the Impact on Plants. **Frontiers in Microbiology**, v. 10, p. 2764, 2019.

TECHNOLOGIES, O. N. **Oxford Nanopore Technologies. Nanopore**. Disponível em: <<https://nanoporetech.com/>>.

THOMAS, J. E. et al. ICTV virus taxonomy profile: Nanoviridae. **Journal of General Virology**, v. 102, n. 3, p. 1–2, 2021.

THOMPSON, J. R. et al. Multiplex RT-PCR detection of four aphid-borne strawberry viruses in *Fragaria* spp. in combination with a plant mRNA specific internal control. **Journal of Virological Methods**, v. 111, p. 85–93, 2003.

TIMCHENKO, T. et al. A Single Rep Protein Initiates Replication of Multiple Genome Components of Faba Bean Necrotic Yellows Virus, a Single-Stranded DNA Virus of Plants. **Journal of Virology**, v. 73, n. 12, p. 10173–10182, 1999.

TIMCHENKO, T.; KATUL, L.; ARONSON, M. Infectivity of nanovirus DNAs : induction of disease by cloned genome components of Faba bean necrotic yellows virus Printed in Great Britain. **Journal of General Virology**, v. 87, p. 1735–1743, 2006.

TIMCHENKO, T.; KATUL, L.; SANO, Y. The Master Rep Concept in Nanovirus Replication : Identification of Missing Genome Components and Potential for Natural Genetic Reassortment. **Virology**, v. 195, p. 189–195, 2000.

VAN DIJK, E. L. et al. Ten years of next-generation sequencing technology. **Trends in Genetics**, v. 30, n. 9, p. 418–426, 2014.

VETTEN, H. J. et al. Nanoviridae. **Elsevier-Academic Press**, p. 342–352, 2005.

VETTEN, H. J. **Encyclopedia of Virology**. 3. ed. Oxford, UK: Elsevier, 2008.

VETTEN, H. J. et al. Identification of a novel nanovirus in parsley. **Archives of Virology**, v. 164, n. 7, p. 1883–1887, 2019.

VIRALZONE. **Nanoviridae**. Disponível em: <https://viralzone.expasy.org/125?outline=all_by_species>.

WANG, Y. et al. Nanopore sequencing technology, bioinformatics and applications. **Nature Biotechnology**, v. 39, n. 11, p. 1348–1365, 2020.

WANITCHAKORN, R., HAFNER, G. J., HARDING, R. M., & DALE, J. L. Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. **Journal of General Virology**, v. 81, p. 299–306, 2000.

ZERBINI, F. M. et al. ICTV virus taxonomy profile: Geminiviridae. **The Journal of General Virology**, v. 98, p. 131–133, 2017.

ZHAO, L. et al. Eukaryotic Circular Rep-Encoding Single-Stranded DNA (CRESS DNA) Viruses : Ubiquitous Viruses With Small Genomes and a Diverse Host Range. **Advances in Virus Research**, v. 103, p. 71–133, 2019.