UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

DEPARTAMENTO DE AGRONOMIA

ÁREA DE FITOSSANIDADE – LABORATÓRIO DE FITOVIROLOGIA

RELATÓRIO DE ESTÁGIO SUPERVISIONADO OBRIGATÓRIO

IDENTIFICAÇÃO E CARACTERIZAÇÃO DE UM NOVO PUTATIVO VÍRUS DA FAMÍLIA *RHABDOVIRIDAE* A PARTIR DE DADOS DE SEQUENCIAMENTO DE ALTO DESEMPENHO DE MANDIOCA (*Manihot esculenta* Crantz)

LUCAS NASCIMENTO DOS SANTOS

RECIFE – PE
2022

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

DEPARTAMENTO DE AGRONOMIA

ÁREA DE FITOSSANIDADE – LABORATÓRIO DE FITOVIROLOGIA

RELATÓRIO DE ESTÁGIO SUPERVISIONADO OBRIGATÓRIO – ESO

IDENTIFICAÇÃO E CARACTERIZAÇÃO DE UM NOVO PUTATIVO VÍRUS DA FAMÍLIA *RHABDOVIRIDAE* A PARTIR DE DADOS DE SEQUENCIAMENTO DE ALTO DESEMPENHO DE MANDIOCA (*Manihot esculenta* Crantz)

Trabalho de Conclusão de Curso apresentado ao Departamento de Agronomia da Universidade Federal Rural de Pernambuco, Unidade Sede, como parte dos requisitos exigidos para obtenção do título de Engenheiro Agrônomo.

Orientadora
Profª Drª. Rosana Blawid

RECIFE – PE
2022

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

DEPARTAMENTO DE AGRONOMIA

ÁREA DE FITOSSANIDADE – LABORATÓRIO DE FITOVIROLOGIA

IDENTIFICAÇÃO

Discente: Lucas Nascimento dos Santos

Curso: Agronomia

Orientadora: Prof.ª Dr.ª Rosana Blawid

Local: Laboratório de Fitovirologia, Departamento de Agronomia - Fitossanidade, Universidade Federal Rural de Pernambuco. Rua Dom Manoel de Medeiros, Dois Irmãos, Recife- PE, 52171-900.

Período: 03/03/2022 a 18/05/2022

Carga horária: 210 horas

AGRADECIMENTOS

Agradeço a Deus, que está presente em todos os momentos de minha vida, que me fez acreditar nos meus sonhos e me abriu portas que eu jamais imaginei, que me concedeu graça e sabedoria para atravessar momentos difíceis em mais uma etapa da vida profissional.

Aos meus pais, Silene Faustina e Antonio Adilson que sempre se sacrificaram e deram o melhor de si para que eu realizasse os meus sonhos e objetivos. Também agradeço ao meu irmão Matheus Nascimento, por toda a ajuda desprendida a mim durante todos estes anos. A todos os familiares, que mesmo um pouco distantes, também foram fundamentais durante a minha trajetória, compreendendo minhas ausências e fornecendo apoio, incentivo e orações.

À Universidade Federal Rural de Pernambuco, pela forma cuidadosa com a qual recebe os seus alunos, aos excelentes professores e demais servidores desta casa.

À professora Rosana Blawid, que de maneira muito gentil aceitou orientar este trabalho, por ser um exemplo de profissionalismo e dedicação ao ensino, pelas oportunidades, confiança e por todo o conhecimento compartilhado.

Ao professor Marco Gama, que é responsável por grande parte do meu crescimento científico durante os quase quatro anos de orientação na Iniciação Científica no LAFIBAC, pela confiança e pelas oportunidades.

Aos muitos amigos e colegas, com os quais tive o prazer de compartilhar bons momentos durante a graduação, na sala de aula ou em diferentes laboratórios de pesquisa, por todo o companheirismo e auxílio.

Aos colegas do Laboratório de Fitovirologia da UFRPE: Géssyka Albuquerque, que acompanhou de perto o desenvolvimento deste trabalho, Marcelo Henrique, Alejandro Risco, Elayni Araújo, Carlos Henrique, Ana Paula e Ailton Cruz.

Aos amigos e colegas que fiz no Laboratório de Fitobacteriologia da UFRPE, que são muitos, mas em especial, agradeço ao Marcelo, Marcelle e Victória pela amizade. Também agradeço ao Me. Pedro Henrique, Me. Bárbara Ribeiro, Dra. Claudeana Souza, Dra. Beatriz Cruz e Dra. Greecy Mirian, pelo conhecimento compartilhado, pelas oportunidades e exemplo de dedicação e persistência na pesquisa, são pesquisadores que em algum momento foram fundamentais para o meu desenvolvimento pessoal e profissional.

LISTA DE ABREVIAÇÕES E SIGLAS

aa	Aminoácidos
CMD	cassava mosaic disease
CMGs	cassava mosaic geminiviruses
CBSD	cassava brown streak disease
CBSVs	cassava brown streak viruses
DNA	Deoxyribonucleic acid
Ha	Hectare
HCN	Ácido cianídrico ou cianeto de hidrogênio
IGJ	Intergenic Junction
Kb	10^3 nucleotídeos
ORF	Open Read Frame
RdRp	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
ssDNA	single-stranded DNA
SDT	Sequence Demarcation Tool
SRA	Sequence Read Archive
t	Tonelada

LISTA DE FIGURAS

LISTA DE TABELAS

RESUMO

A mandioca (*Manihot esculenta* Crantz.) é uma das culturas agrícolas mais importantes no mundo. Apresenta rusticidade, adaptabilidade e múltiplas aptidões na alimentação humana e animal, *in natura* ou após processamento industrial. Dentre os fatores limitantes à cultura da mandioca estão as doenças, sendo possível encontrar uma série de patologias causadas por diferentes organismos, dentre os quais, os vírus estão entre os principais. Para que técnicas de manejo de doenças possam ser estudadas e implementadas no campo, é necessário que o agente causal da doença seja devidamente identificado, e, neste sentido, muitas ferramentas de bioinformática têm auxiliado os estudos taxonômicos e têm ajudado a compreender a diversidade biológica que há dentro de diferentes organismos. Deste modo, este trabalho objetivou realizar a identificação e caracterização *in silico* de sequências virais em dados de sequenciamento de alto desempenho (*High-throughput sequencing*, HTS) de mandioca disponíveis no NCBI. Os arquivos brutos de sequenciamento de mandioca foram obtidos do repositório SRA do NCBI. Os *reads* foram trimados no Trimmomatic v.0.36 e analisados quanto a qualidade com a ferramenta FastQC v.0.11.9. A montagem do genoma foi realizada através do SPAdes v.3.15 (k21, 33,55, 77, 99; --careful), e os *contigs* foram identificados através da ferramenta tBlastX. Os contigs foram estendidos através de mapeamentos consecutivos utilizando a ferramenta BBmap implementada dentro do software Geneious v.R11. O genoma completo foi caracterizado com base em motivos descritos na literatura (espaçadores intergênicos, motivos para início da transcrição e sinais de poliadenilação), além de regiões repetitivas. As ORFs foram identificadas através de análises realizadas com os bancos de dados do UniProt e Pfam. Além disso, foram identificados motivos conservados na proteína L (RdRp) através do alinhamento com sequências de outros gêneros dentro da subfamília *Betarhabdovirinae*. Também foram determinados possíveis sítios de glicosilação dentro da glicoproteína (G) utilizando a ferramenta NetNGlyc 1.0. O software SDTv1.2 foi utilizado para o alinhamento global de sequências do gene L (nt) de todos os representantes da subfamília *Betarhabdovirinae*. A pipeline utilizada neste trabalho foi eficiente para a montagem e caracterização da nova sequência viral encontrada nos dados de HTS de mandioca. O genoma viral montado apresenta um genoma estruturalmente característico à família *Rhabdoviridae*, porém, com a presença de uma ORF extra anterior ao gene L. Com base nos resultados obtidos, provavelmente, trata-se de uma nova espécie de um novo gênero dentro da família *Rhabdoviridae*.

Estudos futuros deverão ser realizados para a confirmação da descoberta desta provável nova espécie viral.

ABSTRACT

Cassava (*Manihot esculenta* Crantz.) is one of the most important agricultural crops in the world. It presents rusticity, adaptability and many different uses for human and animal food, *in natura* or after industrial processing. Among the limiting factors for cassava cultivation are diseases, and it is possible to find a series of pathologies caused by different organisms, among which viruses are the main ones. To disease management techniques be studied and implemented in the field, it is necessary to properly identify the causal agent of diseases. Therefore, many bioinformatics tools have been developed to help taxonomic studies to understand the biological diversity that exists within the different organisms. Thus, this work aimed to identify and characterize viral sequences from High-throughput sequencing (HTS) data of cassava available at the NCBI. First, the cassava sequencing raw files were obtained from the NCBI's SRA repository. The available reads were trimmed with Trimmomatic v.0.36 and analyzed for quality with the FastQC v.0.11.9 tool. Genome assembly was performed using SPAdes v.3.15 (k21, 33,55, 77, 99; --careful), and contigs matches were identified using the tBlastX tool. The chosen contigs were extended by consecutively mapping reads with the BBmap tool implemented in the Geneious v.R11 software. The entire genome was characterized based on motifs already described in the literature (intergenic spacers, transcription initiation motifs and polyadenylation signals), in addition to repetitive regions. The ORFs were identified using the UniProt and Pfam databases. Conservative motifs in the L protein (RdRp) were determined through sequence alignment of viruses from other genera within the *Betarhabdovirinae* subfamily. Possible glycosylation sites within the glycoprotein (G) were also determined by NetNGlyc 1.0. The SDTv1.2 software was used for global sequence alignments of the L gene of all representatives of the subfamily *Betarhabdovirinae*. The pipeline used in this work was efficient for the assembly and characterization of the putative novel virus found by analyzing cassava HTS data. The new assembled viral genome presents a genomic structure characteristic of viruses from the *Rhabdoviridae* family, however, presenting an extra ORF before the L gene. Based on our results, it is probably that the new assembled viral sequence represents a new species within a new genus in the *Rhabdoviridae* family. Future studies should be performed in order to confirm the discovery of this putative new viral sequence.

SUMÁRIO

1. INTRODUÇÃO GERAL

1.1 Aspectos gerais da cultura da mandioca

A mandioca (*Manihot esculenta* Crantz.) é uma planta arbustiva (ALVES, 2001) pertencente a classe das Eudicotiledôneas, ordem Malphighiales e família Euphorbiaceae, sendo uma das primeiras plantas a serem cultivadas no Brasil (CEBALLOS, 2002). Apresenta uma elevada diversidade genotípica e fenotípica (LEKHA et al., 2011; BOAKYE et al., 2013), o que provavelmente lhe concedeu vantagem adaptativa ao longo de sua evolução.

A família Euphorbiaceae encontra-se entre as cinco famílias maiores do reino vegetal (LASTRA; RENAO, 2009) e abriga diferentes gêneros de importância econômica para o Brasil, que podem apresentar múltiplas aptidões, a exemplo da seringueira (*Hevea* spp.) e a mamona (*Ricinus comunis*). Os representantes desta família estão amplamente distribuídos no Brasil, presentes em diferentes tipos de vegetação e condições edafoclimáticas (SÁTIRO et al., 2008). Dentro da ordem Malphighiales, a família Euphorbiaceae é uma das mais complexas, com uma grande diversidade morfológica (FÉLIX-SILVA et al., 2018), e uma das mais antigas, com aproximadamente 60 milhões de anos (OPENSHAW, 2000). Uma característica principal das plantas desta família é a presença de um sistema laticífero, ou seja, há a exsudação de látex quando os ferimentos causados na casca das plantas atingem os dutos laticíferos, onde o látex é produzido (CEBALLOS, 2002).

A planta de mandioca pode atingir quatro metros de altura, seus ramos são cilíndricos, lenhosos e intercalam-se entre nós e entrenós; as folhas são simples e lobadas (possuem de três a nove lóbulos, excepcionalmente onze), constituídas pelo pecíolo e a lâmina foliar (ROGERS; FLEMING, 1973; ALVES, 2001; TOMICH et al., 2008;). Como um produto, as folhas da mandioca possuem elevados teores de proteína bruta, vitaminas, carotenoides e minerais. A sua constituição mineral varia em função da cultivar e das condições climáticas (AWOYINKA et al., 1995).

A mandioca é uma planta monóica, com flores masculinas e femininas em um mesmo indivíduo. Os ramos de onde saem as inflorescências estão localizados no ápice da planta, porém, também podem ser encontrados nas axilas das folhas (também na porção superior da planta). As flores femininas (posicionadas abaixo das flores masculinas) encontram-se em menor quantidade que as flores masculinas. Em uma

mesma inflorescência, as flores femininas abrem entre uma e duas semanas antes das flores masculinas (protoginia), porém, considerando diferentes ramos, as flores masculinas e femininas da planta podem abrir ao mesmo tempo. (ALVES, 2002; JENNINGS; IGLESIAS, 2002). A polinização, quando ocorre, é realizada por insetos (ALVES, 2002).

A propagação da mandioca pode ser realizada através do corte de seus ramos (propagação vegetativa) ou através das sementes botânicas (ALVES, 2002). Embora incomum, a utilização das sementes botânicas na propagação tem um papel fundamental para o melhoramento genético da cultura, pois a variabilidade genética necessária para a seleção nas progênies só pode ser obtida através dos cruzamentos entre diferentes indivíduos (RAJENDRAN et al., 2000).

A raiz da mandioca é o principal órgão da planta em termos industriais e comerciais (LI et al. 2017). Em termos fisiológicos, as raízes são o principal órgão de reserva da planta e, ao contrário do que se pensa, não são raízes tuberosas (como o inhame), mas sim raízes verdadeiras, logo, não podem ser utilizadas para a propagação vegetativa. As raízes são compostas por três tecidos distintos: periderme, córtex (casca) e parênquima (porção comestível) (EKANAYAKE et al., 1997; ALVES 2002). De maneira geral, uma única planta produz de quatro a oito raízes e o número varia em função da variedade, com algumas variedades podendo produzir 20 raízes ou mais. O formato das raízes também pode variar em função do ambiente e do manejo, além da variação que há, naturalmente, entre os diferentes genótipos, existindo quatro formas principais (cilíndrica, cônica, fusiforme e cilíndrico-cônica) (EKANAYAKE et al., 1997). As raízes são excelentes fontes de carboidratos, que compõem cerca de 35% do seu peso fresco (MONTAGNAC et al., 2009). Fibras, lipídeos, proteínas, vitaminas e minerais também estão presentes em diferentes concentrações (DIALLO et al., 2013).

A mandioca é uma planta cianogênica, ou seja, possui glicosídeos cianogênicos em sua composição (TOKARNIA et al., 2002). No Brasil, trata-se da principal planta cianogênica dentre as milhares de espécies presentes no país (RIET-CORREA et al., 2009; RIET-CORREA et al., 2011; TOKARNIA et al., 2012). Dentro do gênero *Manihot*, podemos encontrar dois grandes grupos comerciais, sendo o teor de glicosídeos cianogênicos o limiar para a classificação das espécies dentro destes dois grupos. Ao longo dos anos diferentes autores propuseram diferentes valores relativos aos teores limite de glicosídeos cianogênicos para a classificação destas plantas em mansas ou bravas (LORENZI et al., 1993; SANCHEZ, 2004). O primeiro grupo, e que exige maior

cautela quanto ao consumo alimentar, abriga espécies que apresentam maior teor de glicosídeos cianogênicos no tecido parenquimático das raízes (polpa) e consequentemente, maior toxicidade (mandiocas bravas) e o segundo que apresenta baixo teor de glicosídeos cianogênicos são chamadas de mandiocas mansas (ELIAS et al., 2004; EMPERAIRE; PERONI, 2007; MCKEY et al., 2010). Estes glicosídeos cianogênicos estão associados ao sistema de defesa da planta (SAUNDERS, 2012), porém, quando presentes no alimento, sofrem hidrólise ácida no trato digestivo, produzindo ácido cianídrico, que é tóxico para animais e humanos (TOKARNIA et al., 2002; CÂMARA; SOTO-BLANCO, 2013). O teor destes precursores de ácido clorídrico (HCN) pode variar em função da variedade, da idade da planta e do órgão vegetal amostrado. No entanto, de maneira geral, encontram-se em maior quantidade nas folhas (SILVA et al., 2004; SAUNDERS, 2012). As variedades bravas favorecem o manejo agrícola, apesar de uma toxicidade, em função da maior tolerância aos estresses abióticos, como acidez e baixa fertilidade do solo, além de apresentarem maior resistência a pragas e doenças (FRASSER, 2010), o que eleva o seu potencial produtivo e reduz os custos com insumos e mão de obra empregáveis da cultura (SOUZA et al., 2013).

A mandioca é uma das culturas agrícolas mais importantes no mundo. Rusticidade e adaptabilidade, sobretudo, às condições de clima e solo, lhe conferem sucesso produtivo. É uma cultura que apresenta múltiplas aptidões: alimentação humana e animal, *in natura* ou após processamento industrial (SOUZA, 2017). Segundo dados da FAO, em 2019, a produção mundial foi de 308.647.230 t e a cultura esteve estabelecida em uma área de 27.827.924 ha. Os principais países produtores foram Nigéria, República Democrática do Congo e Tailândia. O Brasil surge na quinta posição em área plantada (5,66%) e produção (4,27%) (FAO, 2021). Apesar dos elevados valores de produção e área plantada, os principais produtores mundiais ainda precisam evoluir em produtividade, a exemplo do Brasil (30ª posição), Nigéria (64ª posição) e República Democrática do Congo (66ª posição) (FAO, 2021).

Dentre os fatores limitantes à cultura da mandioca estão as pragas e doenças, de modo que é possível encontrar uma série de doenças causadas por diferentes organismos, dentre os quais, fungos, bactérias, nematoides e vírus são os principais. Os sintomas das diferentes doenças podem se manifestar em todos os órgãos da planta e quando em condições favoráveis, podem alcançar níveis epidêmicos (KIMATI et al., 1997). Dentre as principais pragas, temos na Ordem Lepidoptera o Mandarová, com ampla distribuição geográfica e com um grande potencial de causar perdas em função de um alto consumo

de área foliar. Além deste, temos representantes importantes em outras ordens, a exemplo da mosca branca, que pode transmitir viroses importantes. Formigas e cupins também são insetos importantes para a cultura, assim como alguns ácaros (GOMES; LEAL, 2003).

1.2 Principais doenças da mandioca causadas por vírus

Diversos vírus foram descritos causando doenças em mandioca ao longo dos anos, na América Latina, destacam-se vírus das famílias *Alphaflexviridae*, *Caulimoviridae* e *Secoviridae*. Na África, destacam-se várias espécies de begomovírus, que causam prejuízos enormes no continente, além dos ipomovírus. Por fim, na Ásia, encontramos principalmente begomovírus e outros vírus não relacionados a epidemias, a exemplo dos ourmiavírus, nepovírus e anulavírus (LEGG et al., 2015).

No continente africano, as doenças de origem viral têm sido o principal problema na cultura da mandioca. Em Benin, Camarões e Nigéria, por exemplo, a Doença do Mosaico da Mandioca (*Cassava Mosaic Disease*, CMD) tem causado sérios problemas no campo, sendo considerada a principal doença destes países (AKINBADE, 2010; ENI et al., 2020; HOUNGUE et al., 2022). A doença é causada por espécies do gênero *Begomovirus* (Família *Geminiviridae*), que são vírus constituídos de uma molécula de ssDNA (DNA de fita simples) circular (ANDRADE; LARANJEIRA, 2019). Os sintomas típicos observados quando há infecção por CMD são mosaico e deformação foliar. A intensidade dos sintomas varia de acordo com a espécie viral (ou espécies, em caso de infecção mista) que estão causando a doença e a susceptibilidade do hospedeiro (NDUNGURU et al., 2005). Atualmente, 11 espécies de begomovírus (**Tabela 4**, material suplementar), também conhecidos como *cassava mosaic geminiviruses* (CMGs) já foram descritos causando CMD (ANDRADE; LARANJEIRA, 2019; LEGG et al., 2015). A mosca-branca (*Bemisia tabaci*) é a principal responsável pela transmissão, mas existem relatos de transmissão através de ferramentas contaminadas durante o corte dos ramos utilizados na propagação vegetativa (FAUQUET et al., 2005).

Ainda na África, dois vírus do gênero *Ipomovirus* (Família *Potyviridae*) também tem causado sérios problemas na cultura da mandioca. São eles: Cassava brown streak virus (CBSV) (WINTER et al., 2010) e Ugandan cassava brown streak virus (UCBSV) (MBANZIBWA et al., 2009). Os sintomas desta doença podem surgir nas folhas, ramos e raízes. Nas raízes, pode ocorrer necrose e constrições radiais; nas folhas, os sintomas são expressos em clorose e mosaico; ocasionalmente, também podem ocorrer lesões ou

estrias de coloração amarronzada (NICHOLS, 1950). Dentro do patossistema do CBSD, podem ocorrer infecções mistas com CBSV e UCBSV (KATHURIMA et al., 2016; MBANZIBWA et al., 2011; OGWOK et al., 2014), porém, pouco se sabe a respeito do efeito destas possíveis interações.

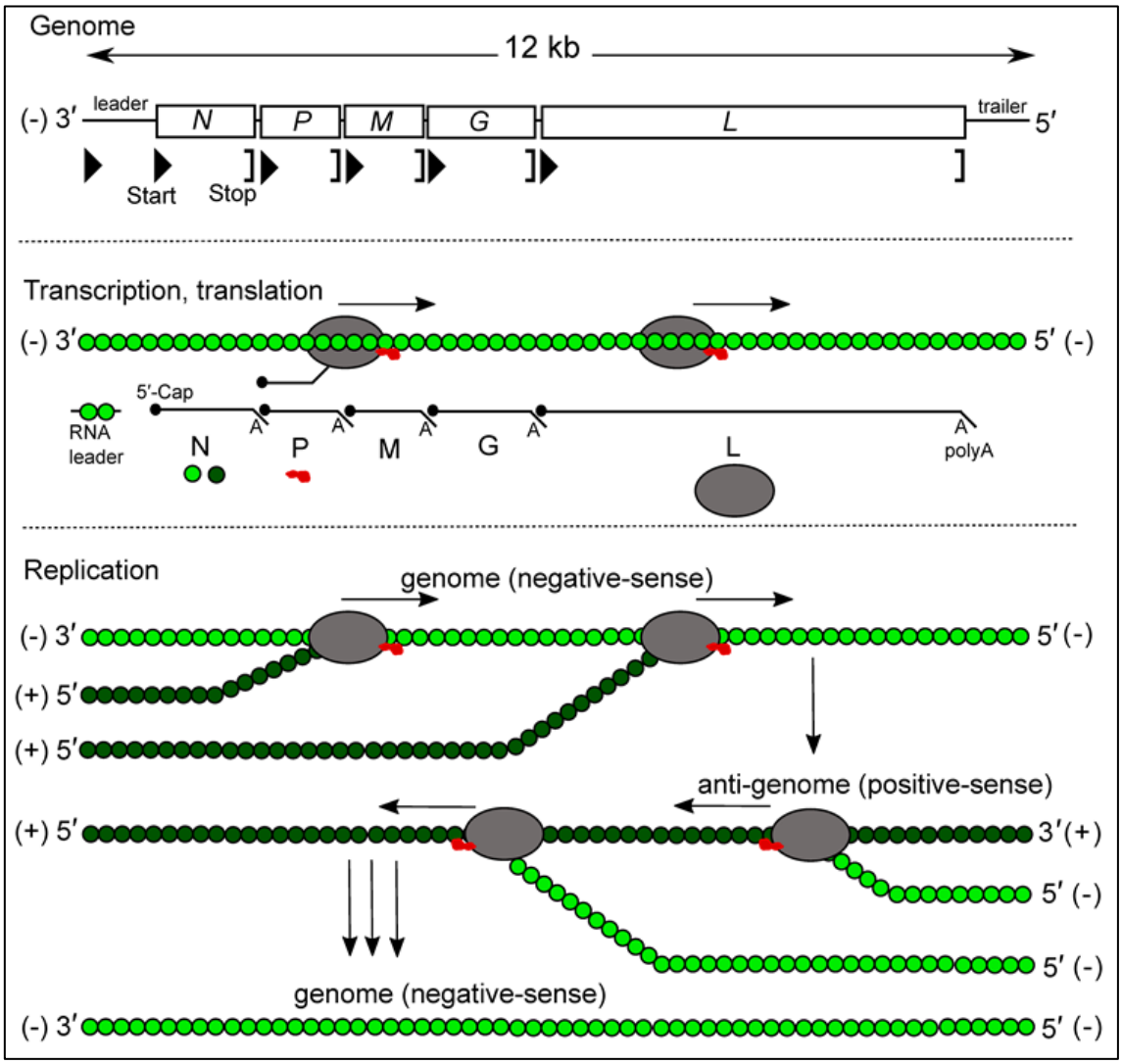
1.3 A família *Rhabdoviridae*

A família *Rhaboviridae* é composta por vírus de RNA fita simples senso negativo, com tamanho que varia entre 10-16 kb. Os vírions podem ser baciliformes ou em formato de bala, envoltos por envelope ou não envelopados. A maior parte destes vírus possui genoma não segmentado, as exceções encontram-se nos gêneros *Dichorhavirus* e *Varicosavirus*, que abrigam vírus de genoma bipartido (WALKER et al., 2021). De maneira geral, os genomas dos rhabdovírus possuem cinco ORFs que codificam, sucessivamente, a partir da extremidade 3' terminal, as proteínas estruturais: nucleoproteína (N), fosfoproteína (P), proteína matriz (M), glicoproteína G e a subunidade maior da RdRp (L) (IVANOV et al., 2011). No entanto, muitas espécies possuem ORFs adicionais entre ou dentro das cinco ORFs canônicas, a exemplo da ORF que codifica a proteína de movimento (MP) em muitos betarhabdovírus, que infectam plantas (WALKER et al., 2021).

O ambiente onde a replicação viral ocorre pode variar nos diferentes rhabdovírus que infectam plantas. Os nucleorhabdovírus são aqueles que replicam no núcleo da célula, a exemplo dos alphanucleorhabdovírus, dichorhavírus e gammanucleorhabdovírus (VAN BEEK et al., 1985; REDINBAUGH et al., 2002; JACKSON et al., 2005; KONDO et al. Al., 2001). Os cytorhabdovírus replicam no citoplasma da célula. Na replicação, há a síntese de uma molécula de RNA de sentido positivo (antigenoma), que é um intermediário para a síntese de uma nova fita do RNA genômico (sentido negativo) (**Figura 1**) (EMERSON; WAGNER, 1972; MOYER et al., 1991)

Os genes dos rhabdovírus são flanqueados com sinais de iniciação e poliadenilação em seu genoma, com um tamanho em torno de 10 nt, o que permite a transcrição de RNAs sub-genômicos (**Figura 1)** (WALKER et al., 2021). Os estudos relativos à compreensão da transcrição e replicação em rhabdovírus ocorreram principalmente a partir dos vesiculavírus e lyssavírus (Subfamília *Alpharhabdovirinae*) (BENERJEE, 1987; FINKE; CONZELMANN, 2005; BENERJEE; BARIK, 1992).

Figura 1. Representação da transcrição e replicação dos rhabdovírus de acordo com Walker et al. (2021)



Atualmente, encontram-se descritas três subfamílias, 45 gêneros e 275 espécies no ICTV (Comitê Internacional de Taxonomia dos Vírus). A subfamília *Alpharhabdovirinae* é a maior e mais diversa subfamília na família *Rhabdoviridae*. Os vírus deste grupo podem infectar uma elevada quantidade de animais, incluindo vertebrados, invertebrados ou vertebrados e seus artrópodes vetores. A segunda maior subfamília (*Betarhabdovirinae*) é composta por vírus que infectam plantas monocotiledôneas e eudicotiledôneas, além de seus vetores (quando conhecidos). *Gammabetarhabdovirinae* é a terceira subfamília, composta apenas pelo gênero *Novirhabdovirus*, cujas espécies encontram-se filogeneticamente distantes dos alpharhabdovírus, e infectam numerosas espécies de teleósteos (peixes ósseos). Há ainda, sete gêneros de vírus que, até então, não foram atribuídos a nenhuma subfamília e estão

relacionados a animais de nichos ecológicos distintos, e, destacam-se neste grupo, os vírus que infectam insetos e ácaros (WALKER et al., 2021).

Os rhabdovírus podem ser responsáveis pela expressão de um grande e variado número de sintomas, que podem se assemelhar aos sintomas causados por outros vírus, de modo que não há poder discriminatório nos sintomas observados. Em virtude deste fator, a microscopia eletrônica demonstrou ser, ao longo dos anos, uma alternativa eficiente para a observação das partículas virais nos tecidos vegetais infectados (WAGNER, 1987).

1.4. Metagenômica e Bioinformática

Durante muitos anos e ainda hoje, a genômica permitiu compreender o conteúdo genético de milhares de organismos em todos os diferentes domínios da vida individualmente. Contudo, nos últimos anos, a metagenômica tem permitido o estudo dos genomas de centenas de organismos de um determinado ambiente em um mesmo tempo, sobretudo, em amostras ambientais (HUGENHOLTZ; TYSON, 2008). É interessante citar, por exemplo, os recentes avanços relativos à compreensão da biodiversidade do solo, de sua complexidade e função através da metagenômica. O solo que é considerado por muitos um dos ambientes mais diversos do planeta (RIESENFELD et al., 2004; DANIEL, 2005). Esta abordagem permite, inclusive, o estudo de organismos de difícil cultivo *in vitro* e até mesmo os não cultiváveis, os quais, compreendem a maior biodiversidade do planeta, com grande potencial biotecnológico. Além disto, a metagenômica permite compreender a dinâmica da comunidade microbiana do solo em função das ações antrópicas (SUYAL et al., 2019).

Em estudos fitopatológicos, a metagenômica tem auxiliado os estudos taxonômicos na identificação e caracterização dos patógenos, especialmente, os vírus de plantas. Através de dados HTS, Santos et al. (2022) puderam estudar a diversidade viral existente em diferentes plantas do gênero *Allium* distribuídas ao redor do mundo (SANTOS et al., 2022). Mendoza et al. (2022) também conseguiram montar genomas virais através de dados HTS de inhame, identificando e caracterizando o yam mosaic virus (YMV) em plantas de *Dioscorea cayennesis-rotundata* (MENDOZA et al., 2022). Embora sejam análises que se complementam, o uso de HTS tem demonstrado, por vezes, uma maior eficiência na detecção viral em amostras vegetais quando comparado com

demais bioensaios, sobretudo, quando se leva em consideração a velocidade e a sensibilidade da análise, que permite, inclusive, a descoberta de vírus ainda não caracterizados (AL RWAHNIH et al., 2015; ROTT et al., 2017; VILLAMOR et al., 2019).

1.5. Critérios para delimitação taxonômica em espécies da família *Rhabdoviridae*

Os critérios para delimitação taxonômica a nível de subfamília são a formação de um clado monofilético a partir de uma árvore filogenética de *Maximum likelihood* (ML) ou árvore filogenética de *Maximum Clade Credibility* (MCC) a partir das sequências completas da proteína L. O uso da proteína L em análises taxonômicas nesta família é justificada pela presença de domínios conservados ao longo da sequência, além da ocorrência incomum de eventos de recombinação. Adicionalmente, são consideradas diferenças significativas na sequência genômica viral e por fim, os aspectos ecológicos, como a gama de hospedeiros (WALKER et al., 2021).

Para delimitação taxonômica a nível de gênero, além dos aspectos considerados na delimitação a nível de subfamília, agora considera-se a arquitetura do genoma, antigenicidade e outras características ecológicas, tais como questões ligadas as patologias e padrões de transmissão (WALKER et al., 2021).

2. MATERIAL E MÉTODOS

2.1. Mineração e download de *Sequence Read Archive* (SRA) e download de HTS. Inicialmente, a mineração dos arquivos brutos de sequenciamento foi realizada manualmente, observando todos os arquivos HTS disponíveis no SRA repositor e realizando seleções com base nas análises taxonômicas preliminares do repositório SRA e da classificação taxonômica obtida com a ferramenta Kaiju (MENZEL et al., 2016). Posteriormente, a mineração destes arquivos brutos de sequenciamento foi realizada com a ferramenta *Serratus* (https://serratus.io/explorer/rdrp) utilizando a função *SRA Run*, através da inserção de um arquivo SRA previamente conhecido contendo *reads* de vírus da família *Rhabdoviridae*, seguindo os parâmetros padrões de identidade de alinhamento e *Score*. Uma planilha de Excel foi gerada e analisada, onde apenas os arquivos SRA de plantas foram escolhidos através da função de formatação condicional (valores duplicados), contrastando todos os números de acesso de arquivos SRA de plantas obtidos no NCBI. Os arquivos promissores foram selecionados manualmente através dos valores de identidade e cobertura observados na planilha.

2.2. Análise de qualidade das sequências e trimagem. A análise de qualidade dos dados de HTS foi realizada consecutivamente com o FastQC v.0.11.9 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), que fornece diferentes gráficos relativos à qualidade dos *reads* de acordo com as notas de Phred (Q) (**Tabela 6**, material suplementar) (PROSDOCIMI et al., 2003), além de indicar a presença de adaptadores. Os cortes das regiões de baixa qualidade e dos adaptadores (trimagem) foram realizados através do Trimmomatic v.0.36 (BOLGER et al., 2014) de acordo com a qualidade observada previamente no FastQC. Os parâmetros de trimagem foram os seguintes: ILLUMINACLIP: TruSeq3-SE (2:30:10); LEADING: 20; TRAILING: 20; SLIDINGWINDOW: 4:20; MINLEN: 36 As etapas, por vezes, foram sucessivamente repetidas até que se obtivesse um arquivo de qualidade satisfatória (Q >20). As trimagens foram realizadas através do servidor *Galaxyuse*, bem como através de linhas de comando na plataforma Linux.

2.3 Montagem dos genomas e mapeamento. Os *reads* foram montados através da estratégia *de novo* com a ajuda do software SPAdes v.3.15 (BANKEVICH et al., 2012) com diferentes valores de k (21, 33, 55, 77 e 99) e utilizando a função *-careful*, que busca realizar uma montagem mais cuidadosa, com um menor número de *mismatches*, ou seja,

um menor número de sítios polimórficos. Os *contigs* foram analisados através da ferramenta tBlastX. Em seguida, os *contigs* com maior similaridade (identidades acima de 70%, e com alta cobertura) a algum representante da família *Rhabdoviridae* foram mapeados consecutivamente com a ferramenta BBmap (BUSHNELL, 2014) implementada no software Geneious v.R11 até a obtenção do genoma viral completo.

2.4 Caracterização do genoma viral. Inicialmente, a caracterização do genoma foi realizada com a ajuda da função *find ORFs* no Geneious que identifica todas as ORFs em todos os quadros de leitura possíveis. As ORFs foram também identificadas através dos bancos de dados do UniProt (The UniProt Consortium, 2017) e Pfam (MISTRY et al., 2021). As regiões intergênicas foram identificadas de acordo com o que já foi descrito na literatura: IGJs (*Intergenic Junction*), que são caracterizadas por um sinal de poliadenilação, espaçador intergênico e sítio de início da transcrição (RAMALHO et al., 2014)

2. 5 Caracterização *in sílico* das proteínas virais. A proteína L foi alinhada com outras sequências L de *Alphanucleorhabdovirus* e *Dichorhavirus* para a determinação dos motivos conservados dentro desta proteína. A glicoproteína (G) foi caracterizada comparativamente de acordo com as informações disponíveis na literatura (COLL, 1995). Foram determinados o tamanho da proteína, o número de resíduos de cisteína e o número de potenciais sítios de N-glicosilação através da ferramenta NetNGlyc 1.0.

2.6 Análise de identidade. O software *Sequence Demarcation Tool* versão 1.2 (SDTv1.2) foi utilizado para o alinhamento global das sequências de genes e proteínas das regiões N, P, M, G e L de todos os representantes da subfamília *Betarhabdovirinae*, com exceção do gênero *Varicosavirus*, que esteve incluso apenas na comparação do gene L.

RESULTADOS E DISCUSSÃO

3.1 Mineração profunda de *Sequence Read Archive* (SRA) e download de HTS. Inicialmente, cerca de 500 arquivos foram analisados manualmente. No entanto, a partir da busca do *Serratus* foi possível identificar 5.076 arquivos brutos de sequenciamento de diferentes organismos, que após a filtragem dos arquivos, tendo como referência os códigos de acesso dos arquivos de plantas, restaram 601 arquivos HTS de plantas. **Na Tabela 1** encontram-se descritas todas as informações relativas aos arquivos com potencial para a montagem de genomas de rhabdovírus de acordo com os 601 arquivos apresentados pelo *Serratus* (aqueles que apresentam valores elevados de cobertura e identidade). Os arquivos selecionados para montagem de genomas virais após a seleção na **Tabela 1** estão descritos na **Tabela 2**.

Tabela 1. Identificação botânica e caracterização dos arquivos SRA estudados através de buscas na plataforma *Serratus*.

Família	Espécie	N*	ID** (%)	C***
Apocynaceae	*Asclepias syriaca (*Algodão bravo)	1	75	961
Asteraceae	*Lactuca sativa (*Alface)	4	98	469
	Lactuca serriola (Alface espinhosa)	2	99	2313
Brassicaceae	*Cardamine amara* (Agrião Grande)	6	78	227
	Cardamine leucantha (Cardamine)	4	82	1586
Ebenaceae	*Diospyros kaki* (Caquizeiro)	76	99	603
Euphorbiaceae	*Manihot esculenta* (Mandioca)	27	65	720
Plantaginaceae	*Bacopa monnieri* (Brahmi)	3	83	53089
Poaceae	*Triticum aestivum* (Trigo)	8	99	36048
Rosaceae	*Prunus persica* (Pêssegueiro)	1	98	1865
Solanaceae	*Lycium barbarum* (Goji Berry)	4	60	695
	Lycium ruthenicum (Goji Berry)	4	50	388
	Solanum lycopersicum (Tomateiro)	4	84	272

*número de arquivos; **identidade e ***cobertura (média)

Tabela 2. Arquivos SRA utilizados na montagem de genomas virais.

Nº de acesso	Espécie	Serratus			GC%**
		score	identidade	cobertura	
DRR168869	*D. kaki*	93	100	1280	46.2
DRR215738	*C. leucantha*	100	83	4246	45.0
SRR10480882	*M. esculenta*	100	67	1799	44.8
SRR10480883	*M. esculenta*	100	66	1692	44.7
SRR10480885	*M. esculenta*	100	65	1505	45.1
SRR10158658	*Nimphaea prolifera*	100	78	11.277	49.9
SRR11473374	*Tr. aestivum*	100	100	4257	56.1
SRR2939237	*S. lycopersicum*	100	84	4346	42.5
SRR5855926	*L. serriola*	100	98	3828	44.4
SRR6705072	*Copstis teeta*	100	79	6114	44.2

*Nº de *reads* **Conteúdo GC (%). SRR10314243 e SRR6705072 foram obtidos através de buscas manuais no NCBI

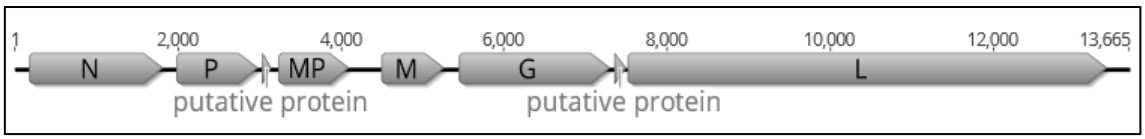
3.2 Análise de qualidade das sequências e trimagem. O FastQC foi eficiente para o detalhamento da qualidade dos *reads* de todos os arquivos trabalhados. A partir da trimagem com o Trimmomatic foi possível remover adaptadores e sequências de baixa qualidade com rapidez e eficiência, sobretudo, quando em linha de comando Linux.

No arquivo utilizado (SRR10480882, **tabela 2**) para a montagem do putativo rhabdovírus não foi necessário realizar cortes no Trimmomatic em função de sua qualidade inicial em nível satisfatório e por não possuir adaptadores em suas sequências. Este arquivo possui 45.522.900 *reads*, com tamanho variando entre 92 e 150 nt (**Figura 9**, material suplementar). O conteúdo GC foi de 44% em média, porém apresentou variação ao longo dos *reads*, de modo que a variação esteve próxima de uma distribuição teórica normal (**Figura 6**, material suplementar). Não houve valores significativos de N, ou seja, posições nas quais não foi possível realizar a identificação da base nitrogenada pelo algoritmo (**Figura 8**, material suplementar). A maior parte dos *reads* possui tamanho superior a 147, o que facilitou a sobreposição dos *reads*, resolvendo possíveis regiões repetitivas e diminuindo a possibilidade de erros na montagem. Não foram encontrados adaptadores dentro dos *reads* (**Figura 10**, material suplementar) e nem sequências super-representadas, a exemplo de *primers*.

3.3 Montagem dos genomas e mapeamento. A montagem do genoma viral foi realizada dentro dos parâmetros descritos anteriormente. O genoma montado com o arquivo SRR1048082 tem um tamanho de 13.665 nt e possui as cinco ORFs canônicas

dos rhabdovírus, além da MP. Todas as ORFs foram identificadas através dos bancos de dados do UniProt e Pfam. Na **Figura 2** é possível observar a estrutura do genoma viral, que apresenta características típicas dos representantes da subfamília *Betarhabdovirinae*, no entanto, com uma ORF extra anterior ao gene L. Esta ORF não está presente em nenhuma espécie até então descrita. Através de mapeamentos com a ferramenta BBmap também foi possível detectar segmentos virais em outros arquivos HTS de mandioca da China (SRR10480883 e SRR10480885) (**Tabela 2**).

Figura 2. Estrutura do genoma montado, mapeado e caracterizado no presente estudo.



Em outras espécies vegetais esta nova putativa sequência viral não foi ainda detectada através de análises por mapeamentos. Deste modo, outros arquivos HTS de mandioca representativos de outras regiões geográficas no mundo devem ser futuramente analisadas para confirmar a hipótese da descoberta de um novo vírus e possível novo gênero dentro da família *Rhabdoviridae*.

3.4. Caracterização *in sílico* das proteínas virais. Através do alinhamento (aa) da região L foi possível identificar quatro motivos conservados dentro da proteína L (**Figura 12**, material suplementar), sendo eles os motivos A, B, C e D, que também já foram identificados em estudos anteriores (RAMALHO et al., 2014; ROY et al., 2015) Através da ferramenta NetNGlyc 1.0 foi possível identificar seis possíveis sítios de N-glicosilação na glicoproteína (G), dentre os quais, dois apresentaram valores de probabilidades elevados na posição 403 (0,7556) e 521 (0,6782) (**Tabela 3**, **Figura 3**). Este resultado está de acordo com o estabelecido em trabalhos anteriores, de acordo com Coll (1995), em que as glicoproteínas de rhabdovírus possuem entre dois e seis potenciais sítios de glicosilação. O tamanho da glicoproteína também está de acordo com o tamanho esperado e possui 16 resíduos de cisteína em sua estrutura. (COLL, 1995).

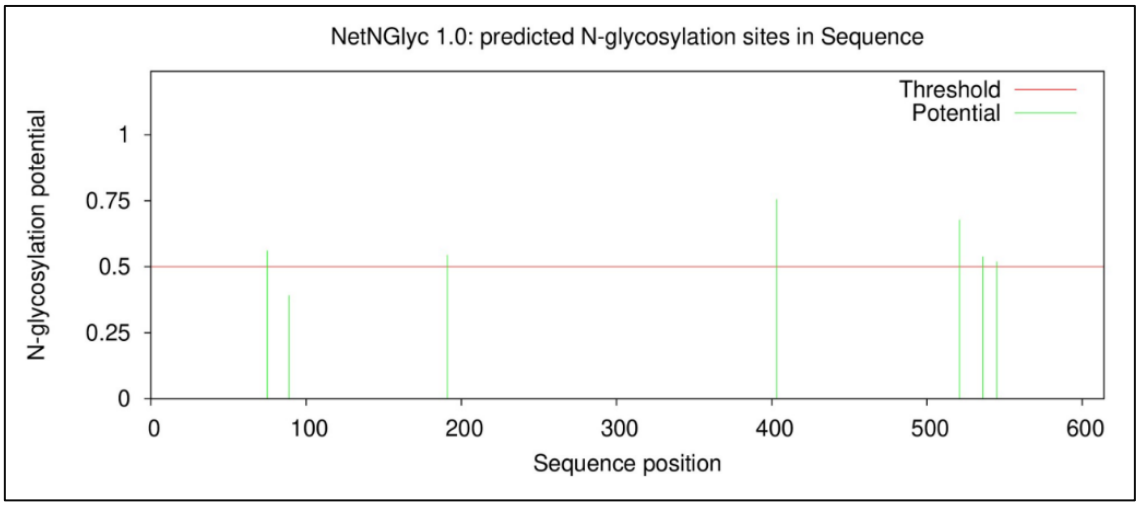
Tabela 3. Possíveis sítios de N-glicosilação identificados no NetNGlyc 1.0.

SeqName	Position	Potential	Juryagreement	N-Glycresult
Sequence 1	75 NSSV	0.5610	(7/9)	+
Sequence 2	89 NSSH	0.3916	(6/9)	-
Sequence 3	191 NISD	0.5442	(6/9)	+

Sequence 4	403 NISV	0.7556	(9/9)	+++
Sequence 5	521 NLTL	0.6782	(9/9)	++
Sequence 6	536 NLSG	0.5384	(7/9)	+
Sequence 7	545 NTTS	0.5191	(5/9)	+

Em vermelho: Sítios com maiores valores de probabilidade de serem glicosilados.

Figura 3. Distribuição dos possíveis sítios de N-glicosilação ao longo da sequência de aminoácidos da glicoproteína (G) identificados através do NetNGlyc 1.0.



3.5 Análise de identidade

A partir da análise do SDT foi possível observar valores de identidade entre 45,5% e 53,8% quando comparamos o gene L a nível de nucleotídeo com todos os representantes da subfamília *Betarhabdovirinae*. A menor identidade observada foi de 45,5% com a sequência L (nt) de orchid fleck virus (gênero *Dichorhavirus*) e as maiores identidades observadas foram de 53,8%, 53,4% e 53,1% com as sequências L de eggplant mottled dwarf virus, physostegia chlorotic mottle virus e potato yellow dwarf virus (gênero *Alphanucleorhabdovirus*), respectivamente. Apesar de observarmos maiores identidade a nível de nt com sequências de alguns alphanucleorhabdovírus será necessário análises adicionais e mais robustas incluindo análises de identidade de outros genes e proteínas virais, análises filogenéticas e caracterização biológica, que visem a confirmação da classificação a nível de gênero da sequência viral montada neste trabalho.

4. CONCLUSÕES GERAIS

O vírus montado no presente trabalho apresenta um elevado potencial de constituir, no futuro, um novo gênero dentro da família *Rhabdoviridae*. Análises futuras ainda precisam ser realizadas para confirmar ou não esta hipótese.

Foi estabelecido um pipeline de montagem de genoma viral desde o tratamento dos dados até a montagem de genomas, podendo ser utilizados em estudos futuros envolvendo rhabdovírus.

Um número de HTS de diferentes origens geográficas ainda devem ser analisados para avaliar se há a presença do vírus montado em diferentes regiões do mundo.

5. REFERÊNCIAS BIBLIOGRÁFICAS

AKINBADE, S. A.; HANNA, R. A.; NGUENKAM, E.; NJUKWE, A.; FOTSO, A.; DOUMTSOP, J.; NGEVE, S. T. N.; TENKU, P.; KUMAR, L. First report of the East African cassava mosaic virus-Uganda (EACMV-UG) infecting cassava (*Manihot esculenta*) in Cameroon. New Disease Reports, vol. 1, issue 1, pag. 22-22, 2010.

AL RWAHNIH, M.; DAUBERT, S.; GOLINO, D.; ISLAS, C.; ROWHANI, A. Comparison of next-generation sequencing versus biological indexing for the optimal detection of viral pathogens of grapevines. Phytopathology 105:758-763, 2015.

ALVES, A. A. C. Chapter Five: Cassava Botany and Physiology IN: HILLOCKS, R. J.; ANANI HOUNGUE, J.; HOUÉDJISSIN, S. S.; AHANHANZO, C.; PITA, J. S.; HOUNDÉNOUKON, M. E.; ZANDJANAKOU-TACHIN, M. Cassava mosaic disease (CMD) in Benin: Incidence, severity and its whitefly abundance from field surveys in 2020. Crop Protection, Volume 158. 2022.

ANDRADE, E. C.; LARANJEIRA, F. F. African cassava mosaic virus (ACMV) e a Doença do Mosaico da Mandioca (Cassava Mosaic Disease, CMD): Subsídios para inclusão de novas espécies de begomovírus causadoras do CMD na lista de pragas quarentenárias A1 do Brasil. EMBRAPA: Documentos 240. 2019.

AWOYINKA, A.F.; ABEGUNDE, V.O.; ADEWUSI, S. R. A. Nutrient content of young cassava leaves and assessment of their acceptance as a green vegetable in Nigeria. Plant Foods Hum Nutr 47(1):21–28. 1995

BABRAHAM BIOINFORMATICS. Fastqc a quality control tool for high throughput sequence data. 2016. Disponível em: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>7

BANKEVICH, A.; NURK, S.; ANTIPOV, D.; GUREVICH, A. A.; DVORKIN M.; KULIKOV, A. S.; LESIN, V. M.; NIKOLENKO, S. I.; PHAM, S.; PRJIBELSKI, A. D.; PYSHKIN, A. V.; SIROTKIN, A. V.; VYAHHI, N.; TESLER, G.; ALEKSEYEV, M. A.; PEVZNER, P. A. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology, vol. 19, No. 5.2012.

BARTHELEMY J. P. LOGNAY, G. Nutritional importance of cassava and perspectives as a staple food in Senegal. A review. Biotechnol Agron Soc Environ. 17:634–643. 2013.

BENERJEE, A.; BARIK, S. Gene expression of vesicular stomatitis virus genome RNA. Virology, 188 (2):447-28, 1992.

BENERJEE, A. K.; Transcription and Replication of rhabdoviruses. Microbiol Rev, 51(2):299, 1987.

BERRIE, L. C., PALMER.; K. E., RYBICKI, E. P.; REY, M. E. C. Molecular characterization of a distinct South African cassava infecting geminivirus. Archives of Virology, v. 143, p. 2253-2260, 1998.

BLAWID, R.; SILVA, J. M. F.; NAGATA, T. Discovering and sequencing new plant viral genomes by next-generation sequencing: description of a practical pipeline. 2017. Annals of Applied Biology, ISSN 0003-4746.

BOAKYE, P. B.; KWADWO, O.; ISAAC, A. K.; PARKES, E. Y. Genetic variability of three cassava traits across three locations in Ghana. African Journal of Plant Science. Vol. 7(7), pp. 265-267. 2013.

BOCK, K. R.; WOODS, R. D. Etiology of African cassava mosaic disease. Plant Disease, v. 67, p994-995, 1983.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics, Volume 30, Issue 15, 1 August 2014, Pages 2114–2120.

BOLGER, A.M.; LOHSE, M.; USADEL, B. Trimmomatic: aflexible trimmer for Illumina sequence data. Bioinformatics, 30, 2114–2120. 2014.

BULL, S. E.; BRIDDON, R. W.; SSERUBOMBWE, W. S.; NGUGI, K.; MARKHAM, P. G.; STANLEY, J. Genetic diversity and phylogeography of cassava mosaic viruses in Kenya. Journal of General Virology, v,87, p.3053-3065, 2006.

BUSHNELL, B. BBMap: A Fast, Accurate, Splice-Aware Aligner. United States: N. p., 2014.

CÂMARA, A. C. L.; SOTO-BLANCO, B. Cyanide poisoning in animals and humans. In: SOTO-BLANCO, B. (Ed.). Cyanide: occurrence, characteristics and applications. Hauppauge: Nova Science Publishers, pag. 23-46. 2013.

CEBALLOS, H. Taxonomia e morfologia de la Yuca. In: OSPINA, I.A.; CEBALLOS, Chemistry in New Zealand. New Zeland, v. 76, n. 4, p. 129-132, 2012.

CHRISTIAN S. RIESENFELD, C. S.; PATRICK D. SCHLOSS; JO HANDELSMAN. Metagenomics: Genomic Analysis of Microbial Communities. Annual Review of Genetics 38:1, 525-552. 2004.

COLL, J. M. The glycoprotein G of rhabdoviruses. Arch Virol. 140(5):827-51. 199.

DANIEL, R. The metagenomics of soil. Nat Rev Microbiol 3, 470–478. 2005.

DIALLO, Y.; GUEYE, M. T.; SAKHO, M.; DARBOUX, P. G., KANE A.
EKANAYAKE, I. J.; OSIRU, D. S. O.; PORTO, M. C. M. Morphology of cassava. Ibadan: International Institute of Tropical Agriculture - IITA, 1997. Guide 61. Disponível em <http://www.iita.org/cms/details/trn_mat/irg61/irg61.html>.

ELIAS, M.; MÜHLEN, G. S.; MCKEY, D.; ROA, A. C.; TOHME, J. Genetic diversity of traditional South American Landraces of Cassava (*Manihot esculenta* Crantz): an analysis using microsatellites. Econ Bot 58, 242–256. 2004)

EMPERAIRE, L.; PERONI, N. Traditional management of agrobiodiversity in Brazil: a case study of manioc. Human Ecology, 35(6), 761-768. 2007.

ENI, A. O.; EFEKEMO, O. P.; ONILE-ERE, O. A.; PITA, J. S. South West and North Central Nigeria: Assessment of cassava mosaic disease and field status of African cassava mosaic virus and East African cassava mosaic virus. Annals of Applied Biology, issue 3, pag. 466-479. 2021.

FAO. FAOSTAT. Crops and livestock products. Latest update:September 15, 2021. Accessed: October 29, 2021. [https://www.fao.org/faostat/en/#data].

FAUQUET, C.; MAYO, M.; MANILOFF, J.; DESSELBERGER, U.; BALL, L. Virus taxonomy, VIIIth Report of the International Committee on Taxonomy of Viruses. Elsevier/Academic, London. 2005

FÉLIX-SILVA, J.; GOMES, J. A. S.; FERNANDES, J. M.; MOURA, A. K. C.; MENEZES, Y. A. S.; SANTOS, E. G. G.; TAMBOURGI, D. V.; SILVA-JUNIOR, A.A.; ZUCOLOTTO, S. M.; FERNANDES-PEDROSA, M. F. Comparison of two *Jatropha* species (Euphorbiaceae) used popularly to treat snakebites in Northeastern Brazil: Chemical profile, inhibitory activity against *Bothrops erythromelas* venom and antibacterial activity. Journal of Ethnopharmacology, vol. 213, pag-12-20 2018

FINKE, S.; CONZELMANN, K. K. Replication strategies of rabies virus. Virus Res. 111(2):120-31. 2005.

FONDONG, V. N.; PITA, J. S.; REY, C.; BEACHY, R. N.; FAUQUET, C. M. First report of the presence of East African cassava mosaic virus in Cameroon. Plant Disease, v.82, p.1172, 1998.
FRASER, J. A. The diversity of bitter manioc (*Manihot esculenta Crantz*) cultivation in a whitewater Amazonian landscape. Diversity, v. 2, n. 4, p. 586-609. 2010.

GOMES, J. C.; LEAL, E. C. Cultivo da Mandioca para a Região dos Tabuleiros Costeiros.Embrapa Mandioca e Fruticultura. Sistemas de Produção, 11. ISSN 1678-8796 Versão eletrônica. Jan, 2003.

GUPTA, R.; JUNG, E.; BRUNAK, S. NetNGlyc 1.0 Server: Prediction of N-glycosylation sites in human proteins. 2004. Disponívelem<http://www.cbs.dtu.dk/services/NetNGlyc/>

H. La Yuca en el tercer milenio. Cali: CIAT, Publicacion. 327, p. 17-33. 2002.

HONG, Y. G.; ROBINSON, D. J., HARRISON, B. D. Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted geminiviruses in cassava. Journal of General Virology, v.74, p.2437–2443, 1993.

HUGENHOLTZ, P.; TYSON, G. W. Metagenomics. Nature, vol 455. 2008.

JACKSON, A. O., DIETZGEN, R. G.; GOODIN, M. M.; BRAGG, J. N; DENG, M. Biology of plant rhabdoviruses. Annu Rev Phytopathol 43: 623-660. 2005.

JACKSON, A. O.; FRANCKI, R. I. B.; ZUIDEMA, D. Biology, Structure, and Replication of Plant Rhabdoviruses. IN: WAGNER, R. R. The rhabdoviruses, springer, 1987.

JENNINGS, D. L.; IGLESIAS, C.A. Breeding for crop improvement. In: HILLOCKS, R. J.; THRESH, J. M.; BELLOTTI, A. C. (Eds.), Cassava: Biology, Production and Utilization. CABI Publishing, pp. 149–166. 2002.

KATHURIMA, T.; NYENDE, A.; KIARIE, S.; ATEKA, E. Genetic diversity and distribution of Cassava brown streak virus and Ugandan cassava brown streak virus in major cassava-growing regions in Kenya. Annu. Res. Rev. Biol. 10, 1–9, 2016.

KIMATI, H.; AMORIM, L.; REZENDE; J. A. M.; FILHO, A. B.; CAMARGO, L. E. A. Manual de Fitopatologia: Doenças de plantas cultivadas. 3. ed. 2 v. São Paulo: Agronômica Ceres, 1997.

KONDO, H.; MAEDA, T.; TAMADA, T. Orchid fleck virus: *Brevipalpus californicus* mite transmission, biological properties and genome structure. Exp Appl Acarol 30: 215-23. 2003.

LASTRA, C. A. M.; HENAO, C. A. A. Taxonomic study of Euphorbiaceae from Quindío (Colombia). Rev. Asoc. Col. Cienc. Biol. (Col.), Vol. 21, pag. 156-173. 2009.

LEGG, J. P.; KUMAR, P. L.; MAKESHKUMAR, T.; TRIPATHI, L.; FERGUSON,

M.; KANJU, E.; NTAWURUHUNGA, P.; CUELLAR, W. Cassava Virus Diseases: Biology, Epidemiology, and Management. Advances in Virus Research. Academic Press, Vol. 91, Pag. 85-142. 2015.

LEKHA, S. S.; SILVA, J. A. T.; PILLAI, S. V. Genetic variability studies between released varieties of cassava and central Kerala cassava collections using SSR markers. Journal of Stored Products and Postharvest Research Vol. 2(4) pp. 79 – 92. 2011.

LI, S. CUI, Y.; ZHOU, Y.; LUO, Z.; LIU, J.; ZHAO, M. The industrial applications of cassava: current status, opportunities and prospects. J Sci Food Agric. 97: 2282–2290. 2017.

LORENZI, J. O.; RAMOS, M. T. B.; MONTEIRO, D. A.; VALLE, T. L.; GODOY

JÚNIOR, G. Teor de ácido cianídrico em variedades de mandioca cultivadas em quintais do Estado de São Paulo. Braganta, Campinas, v. 52, p. 1-5, 1993.

MARUTHI, M.; SEAL, S.; COLVIN, J.; BRIDDON, R. W.; BULL, S. E. East African cassava mosaic Zanzibar virus – a recombinant begomovirus species with a mild phenotype. Archives of Virology, v.149, p.2365-2377, 2004.

MBANZIBWA, D. R.; TIAN, Y. P.; TUGUME, A. K.; PATIL, B. L.; YADAV, J. S.; BAGEWADI, B.; ABARSHI, M. M.; ALICAI, T.; CHANGADEYA, W.; MKUMBIRA, J.; MULI, M. B.; MUKASA, S. B.; TAIRO, F.; BAGUMA, Y.; KYAMANYWA, S.; KULLAYA, A.; MARUTHI, M. N.; FAUQUET, C. M.; VALKONEN, J. P. T. Evolution of Cassava brown streak disease-associated viruses. J. Gen. Virol. 92, 974–987, 2011.

MBANZIBWA, D. R.; TIAN, Y.; MUKASA, S. B.; VALKONEN, J. P. Cassava brown streak virus (Potyviridae) encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contains no HC-PRO. Journal of Virology, 83, 6934-6940, 2010.

MCKEY, D.; CAVAGNARO, T. R.; CLIFF, J.; GLEADOW, R. Chemical ecology in coupled human and natural systems: people, manioc, multitrophic interactions and global change. Chemoecology 20, 109–133. 2010.

MENDOZA, A. R.; MARGARIA, P.; NAGATA, T.; WINTER, S. BLAWID, R. Characterization of yam mosaic viruses from Brazil reveals a new phylogenetic group and possible incursion from the African continent. Virus Genes, 2022.

MENZEL, P.; NG, K. L.; KROGH, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. 2016. Nat. Commun. 7:11257.

MENZEL, P.; NG, K.L.; KROGH, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat. Commun. 7:11257. 2016.

MISTRY, J.; SARA CHUGURANSKY, LOWRI WILLIAMS, MATLOOB QURESHI, GUSTAVO A SALAZAR, ERIK LL SONNHAMMER, SILVIO CE TOSATTO, LISANNA PALADIN, SHRIYA RAJ, LORNA J RICHARDSON, ROBERT D FINN, ALEX BATEMAN, Pfam: The protein family database in 2021, Nucleic Acids Research, Volume 49, Ed. D1, Pag. D412–D419. 2021

MONTAGNAC, J. A.; DAVIS, C. R.; TANUMIHARDJO, S. A. Nutritional value of cassava for use as a staple food and recent advances for improvement. Compr Rev Food Sci Food Saf 8:181–194. 2009.

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

NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research, 2016. 44(Database issue): D7–D19. doi: 10.1093/nar/gkv1290.

NDUNGURU, J.; LEGG, J. P.; AVELING, T. A. S.; THOMPSON, G.; FAUQUET, C. M. Molecular biodiversity of cassava begomoviruses in Tanzania: evolution of cassava geminiviruses in Africa and evidence for East Africa being a center of diversity of cassava geminiviruses. Virology Journal, v.2, p. 21, 2005

NICHOLS, R. F. W. The brown streak disease of cassava. East Afr. Agric. J. 15, 154–160, 1950

OGWOK, E.; ALICAI, T.; REY, M. E. C.; BEYENE, G.; TAYLOR, N. J. Distribution and accumulation of Cassava brown streak viruses within infected cassava (*Manihot esculenta* Crantz) plants. Plant Pathol. 64, 1235–1246. ,2014.

OPENSHAW, K. A review of *Jatropha curcas*: An oil plant of unfulfilled promise. Biomass and Bioenergy, vol. 19 (1), pag. 1-15. 2000.

PROSDOCIMI, F.; PEIXOTO, F. C.; ORTEGA, J. M. DNA SEQUENCES BASE CALLING BY PHRED: ERROR PATTERN ANALYSIS. R. TECNOL. INF. BRASÍLIA V. 3 N. 2 p. 107-110, 2003.

RAMALHO, T. O.; FIGUEIRA, A. R.; SOTERO, A. J.; WANG, R.; DUARTE, P. S. G.; FARMAN, M.; GOODIN, M. M. Characterization of Coffee ringspot virus-Lavras: A model for an emerging threat to coffee production and quality. Virology. Vol. 464–465, Pag. 385-396. 2014.

REDINBAUGH, M. G.; SEIFERS, D. L.; MEULIA, T.; ABT, J. J.; ANDERSON, R. J.; STYER, W. E.; ACKERMAN, J.; SALOMON, R.; HOUGHTON, W.; CREAMER, R.; GORDON, D. T.; HOGENHOUT, S. A. Maize fine streak virus, a new leafhopper-transmitted rhabdovirus. Phytopathology 92: 1167-1174, 2002.

RIET-CORREA, F.; BEZERRA, C. W. C.; MEDEIROS, R. M. T. Plantas tóxicas do Nordeste. Patos: Sociedade Vicente Palloti, 2011. 82 p.

RIET-CORREA, F.; MEDEIROS, R. M. T.; PFISTER, J.; SCHILD, A. L.; DANTAS, A. F. M. Cyanogenic plants. Poisonings by plants, mycotoxins and related substances in brazilian livestock. Patos: Sociedade Vicente Palloti, pag. 149-157.2009.

ROGERS, D. J.; FLEMING, H. S. Monograph of *Manihot esculenta* Crantz. Economic Botany 27, 1–114. 1973.

ROTT, M.; XIANG, Y.; BOYES, I.; BELTON, M.; SAEED, H.; KESANAKURTI, P.; HAYES, S.; LAWRENCE, T.; BIRCH, C.; BHAGWAT, B.; RASI, H. Application of next generation sequencing for diagnostic testing of fruit tree viruses and viroids. Plant Dis. 101:1489-1499, 2017.

ROY, A.; STONE, A. L.; SHAO, J.; OTERO-COLINA, G.; WEI, G.; CHOUDHARY, N.; ACHOR, D.; LEVY, L.; NAKHLA, M. K.; HARTUNG, J. S.; SCHNEIDER, W. L.; BRLANSKY, R. H. Identification and Molecular Characterization of Nuclear Citrus leprosis virus,a Member of the Proposed Dichorhavirus Genus Infecting Multiple Citrus Species in Mexico. Virology, Vol. 105, No. 4, 2015.

SANCHEZ, T. Evaluacion de 6000 variedades de Iuca. Cali, Colômbia: CIAT/Centro Internacional de Agricultura Tropical, 2004. (Programa de mejoramiento de Iuca).

SANTOS, J. A. C. M.; SILVA, A. M. F.; DE MESQUITA, J. C. P.; BLAWID, R. Transcriptomic analyses reveal highly conserved plant amalgavirus genomes in different species of Allium. Acta Virologica. 66(1):11-17, 2022.

SÁTIRO, L. N.; ROQUE, N. A família Euphorbiaceae nas caatingas arenosas do médio rio São Francisco, BA, Brasil. Acta Botânica Brasilica, Vol. 22 (1), pag. 99-118. 2008

SAUNDERS, D. A. When plants bite back: A broadly applicable method for the determination of cyanogenic glycosides as hydrogen cyanide in plant-based foodstuffs.

SAUNDERS, K.; SALIM, N.; MALIC, V. R.; MALATHID, V. G.; BRIDDON, R.; MARKHAM, P. G.; STANLEY, J. Characterization of Sri Lankan Cassava Mosaic

Virus and Indian Cassava Mosaic Virus: Evidence for Acquisition of a DNA B Component by a Monopartite Begomovirus. Virology, v.293, p.63-74, 2002.

SILVA, G. G. C.; NUNES, C. G. F.; OLIVEIRA, E. M. M.; SANTOS, M. A. Toxicidade cianogênica em partes da planta de cultivares de mandioca cultivados em Mossoró-RN. v. 51, n. 293. 2004.

SOUZA, R. F.; SILVA, I. F.; SILVEIRA, F. P. M.; NETO, M. A. D.; ROCHA. I. T. M. Análise econômica no cultivo da mandioca. Revista Verde de Agroecologia e Desenvolvimento Sustentável, v. 8, n. 2, p. 141-150, 11. 2013

SOUZA, R. G. MANDIOCA: RAIZ, FARINHA E FÉCULA: Conjuntura mensal. CONAB, Companhia Nacional de Abastecimento. Fevereiro de 2017.

SUYAL, D. C.; JOSHI, D.; DEBBARMA, P.; SONI, R.; DAS, B.; GOEL, R. Soil Metagenomics: Unculturable Microbial Diversityand Its Function. In: Varma, A., Choudhary, D. (eds) Mycorrhizosphere and Pedogenesis. Springer, Singapore. 2019.

SWANSON, M. M.; HARRISON, B. D. Properties, relationships and distribution of cassava mosaic geminiviruses. Tropical Science, v.34, P.15–25, 1994.

TAMURA, K.; STECHER, G.; KUMAR, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and evolution. Vol. 38, Ed. 7, Pag. 3022–3027. 2021.

THE UNIPROT CONSORTIUM. UniProt: a hub for protein information. Nucleic Acids Research, Vol. 43, Issue D1, Pag. D204–D212. 28 Jan. 2015.

THRESH, J. M.; BELLOTTI, A. Cassava : biology, production, and utilization. CABI Publishing. 2001.

TIENDRÉBÉOGO, F.; LEFEUVRE, P.; HOAREAU, M.; HARIMALALA, M. A.; BRUYN, A.; VILLEMOT, J.; TRAORÉ, V. S. E.; KONATÉ, G.; TRAORÉ, A. S.; BARRO, N.; REYNAUD, B.; TRAORÉ, O.; LETT, J. M. Evolution of African cassava mosaic virus by recombination between bipartite and monopartite begomoviruses. Virology Journal, v.9, p.67, 2012.

TOKARNIA, C. H.; BRITO, M. F.; BARBOSA, J. D.; VARGAS, P. V.; DÖBEREINER, J. Plantas cianogênicas. Plantas tóxicas do Brasil para animais de produção. 2. ed. Rio de Janeiro: Helianthus, p. 443-460. 2012.

TOKARNIA, C. H.; DÖBEREINER, J.; VARGAS, P. V. Poisonous plants affecting livestock in Brazil. Toxicon, Oxford, v. 40, n. 12, p. 1635-1660. 2002.

TOMICH, R. G. P.; SALIS, S. M.; FEIDEN, A.; CURADO, F. F.; SANTOS, G. G.;

TOMICH, T. R. Etnovariedades de mandioca (*Manihot esculenta* Crantz) cultivadas em assentamentos rurais, MS. Dados eletrônicos. - Corumbá: Embrapa Pantanal. 27 p. (Boletim de Pesquisa e Desenvolvimento/ Embrapa Pantanal, ISSN 1981-7215; 78) 2008.

VAN BEEK, N. A. M., LOHUIS, D.; DIJKSTRA J.; PETERS, D. Morphogenesis of sonchus yellow net virus in cowpea protoplasts. J Ultrastr Res 90: 294-303. 1985

VANOV, I.; YABUKARSKI, F.; RUIGROK, R.W. H.; JAMIN, M. Structural insights into the rhabdovirus transcription/replication complex. Virus Research. Vol. 162, Issues 1–2, Pag. 126-137. Dez 2011.

VILLAMOR, D. E. V.; HO, T.; AL RWAHNIH, M.; MARTIN, R. R.; TZANETAKIS, I. E. High Throughput Sequencing For Plant Virus Detection and Discovery. Phytopathology. 109(5):716-725, 2019.

WALKER, P. J.; FREITAS-ASTÚA, J.; BEJERMAN, N.; BLASDELL, K. R.; BREYTA, R.; DIETZGEN, R. G.; FOOKS, A. R.; KONDO, H.; KURATH, G.; KUZMIN, I. V.; RAMOS-GONZÁLEZ, P. L.; SHI, M.; STONE, D. M.; TESH, R. B.; TORDO, N.; VASILAKIS, I.; WHITFIELD, A. E. ICTV Virus Taxonomy Profile: Rhabdoviridae 2021, Journal of General Virology. 2021.

WALKER, P. J.; BLASDELL, K. R.; CALISHER, C. H.; DIETZGEN, R. G.; KONDO, H.; KURATH, G.; LONGDON, B.; STONE, D. M.; TESH, R. B.; TORDO, N.; VASILAKIS, N.; WHITFIELD, A. E.; and ICTV Report Consortium. ICTV Virus Taxonomy Profile: Rhabdoviridae, Journal of General Virology, 99:447–448. 2018.

WINTER, S.; KOERBLER, M.; STEIN, B.; PIETRUSZKA, A.; PAAPE, M.; BUTGEREITT, A. Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. Journal of General Virology, vol. 91, issue 5, 2010.

ZHOU, X.; ROBINSON, D. J.; HARRISON, B. D. Types of variation in DNA-A among isolates of East African cassava mosaic virus from Kenya. Malawi and Tanzania. Journal of General Virology, v.79, p.2835-2840, 1998.

6. MATERIAL SUPLEMENTAR

Tabela 4. CMGs causadores de CMD na África de acordo com Andrade e Laranjeira (2019)

Espécie	Referência
African cassava mosaic virus (ACMV)	BOCK & WOODS, 1983
African cassava mosaic Burkina Faso virus (ACMBFV)	TIENDRÉBÉOGO et al., 2012
Cassava mosaic Madagascar virus (CMMGV)	TIENDRÉBÉOGO et al., 2012
East African cassava mosaic virus (EACMV)	SWANSON & HARRISON, 1994
East African cassava mosaic Cameroon virus (EACMCV)	FONDONG et al., 1998
East African cassava mosaic Kenya virus (EACMKV)	BULL et al., 2006
East African cassava mosaic Malawi virus (EACMMV)	ZHOU et al., 1998
East African cassava mosaic Zanzibar virus (EACMZV)	MARUTHI et al., 2004
South African cassava mosaic virus (SACMV)	BERRIE et al., 1998
Indian cassava mosaic virus (ICMV)	HONG et al., 1993
Sri Lankan cassava mosaic virus (SLCMV)	SAUNDERS et al., 2002

Tabela 5. CBSVs causadores de CBSD na África (LEGG et al., 2015)

Espécie	Referência
Cassava brown streak virus (CBSV)	WINTER et al., 2010
Ugandan cassava brown streak virus (UCBSV)	MBANZIBWA et al., 2009

Tabela 6. Notas de Phred com respectivas probabilidades de erro (Pe) e segurança.

Notas de Phred	Pe	Segurança
10	1 em 10	90%
20	1 em 100	99%
30	1 em 1.000	99,9%
40	1 em 10.000	90,99%
50	1 em 100.000	90,999%

Figura 4. Distribuição da qualidade (Notas de Phred) de acordo com cada posição no *read*. (SRR10480882).

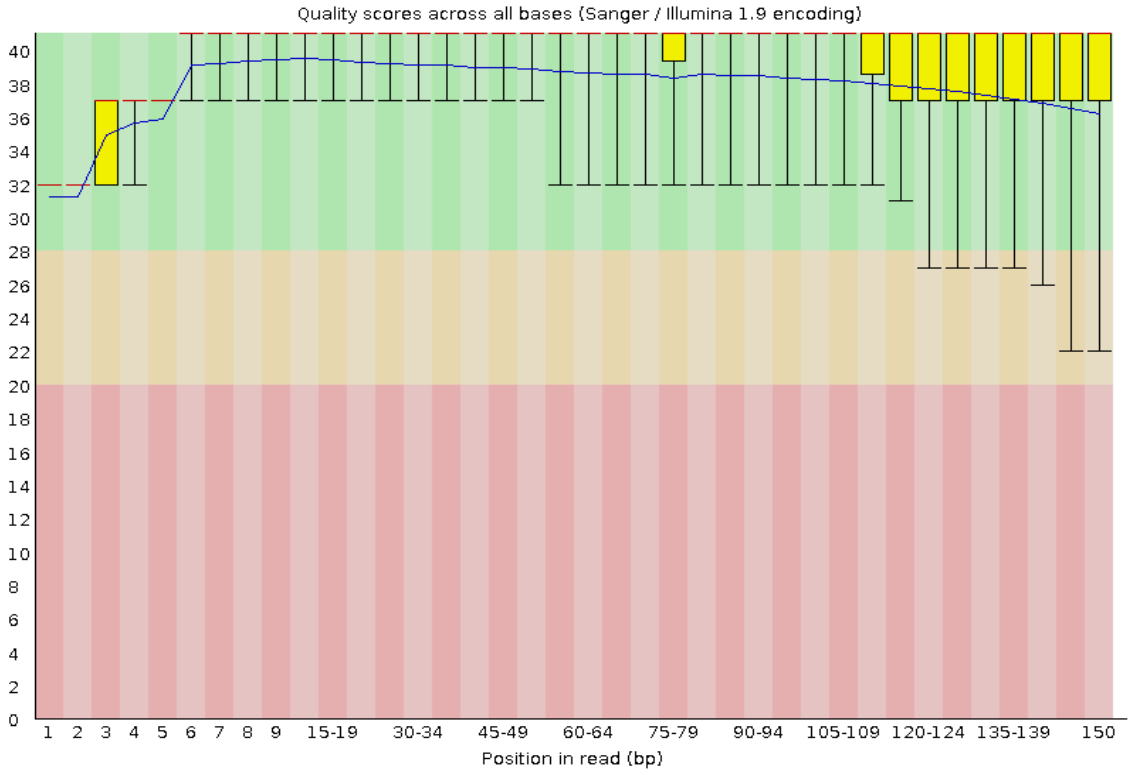


Figura 5. Quantidade dos *reads* de acordo com qualidade

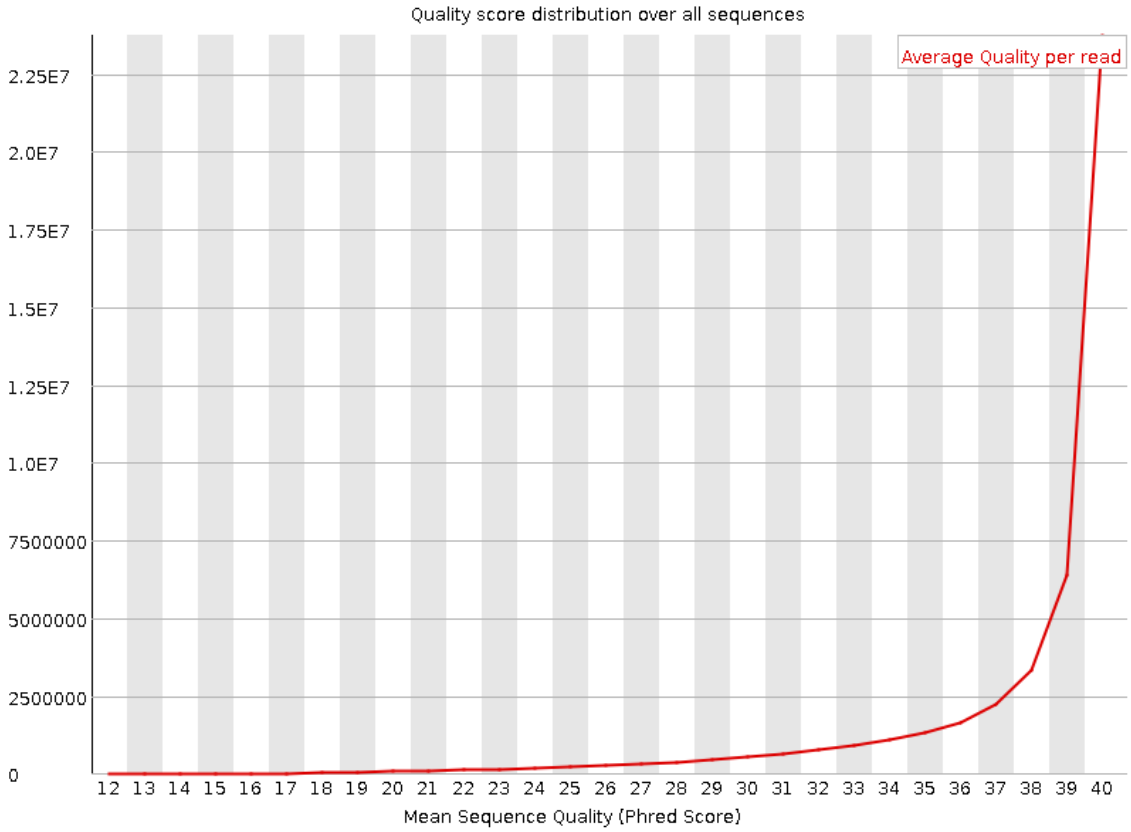


Figura 6. Proporção de cada base nitrogenada em função da posição nos *reads*.

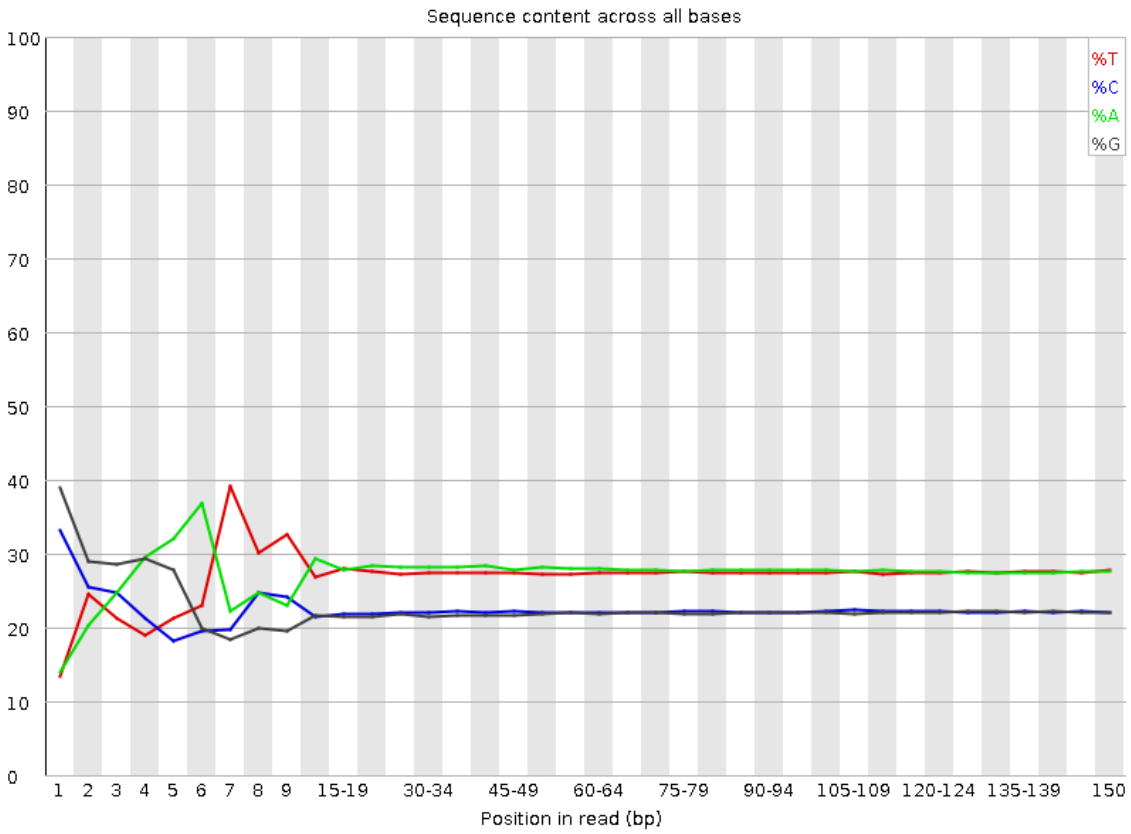


Figura 7. Distribuição do conteúdo GC (%) quando todos os *reads* são considerados

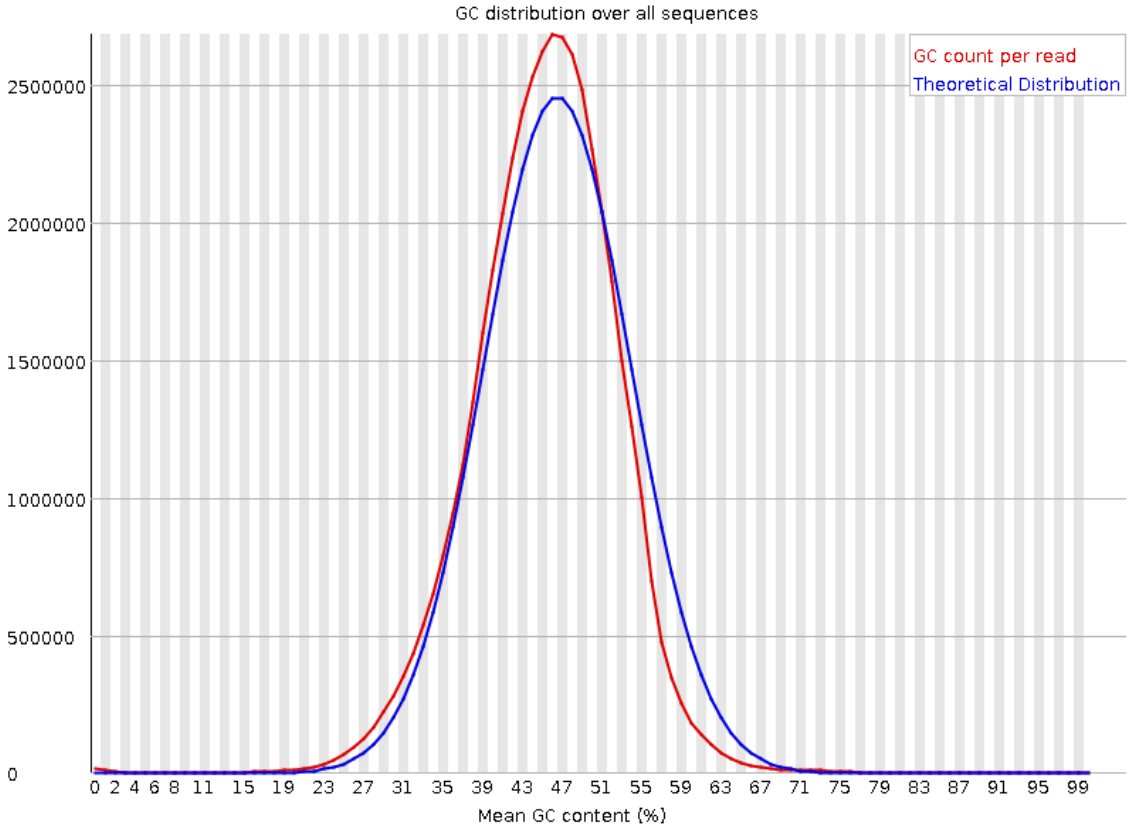


Figura 8. Número de N em função da posição no *reads*

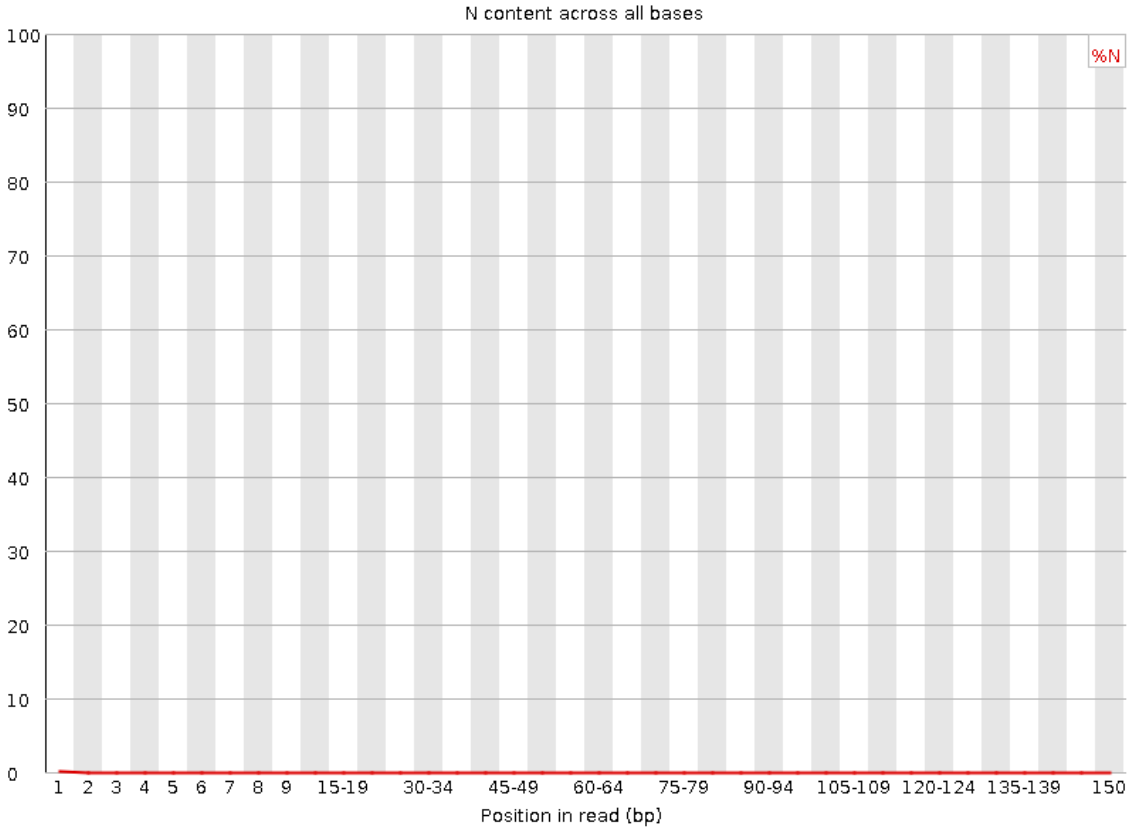


Figura 9. Distribuição do tamanho dos *reads*

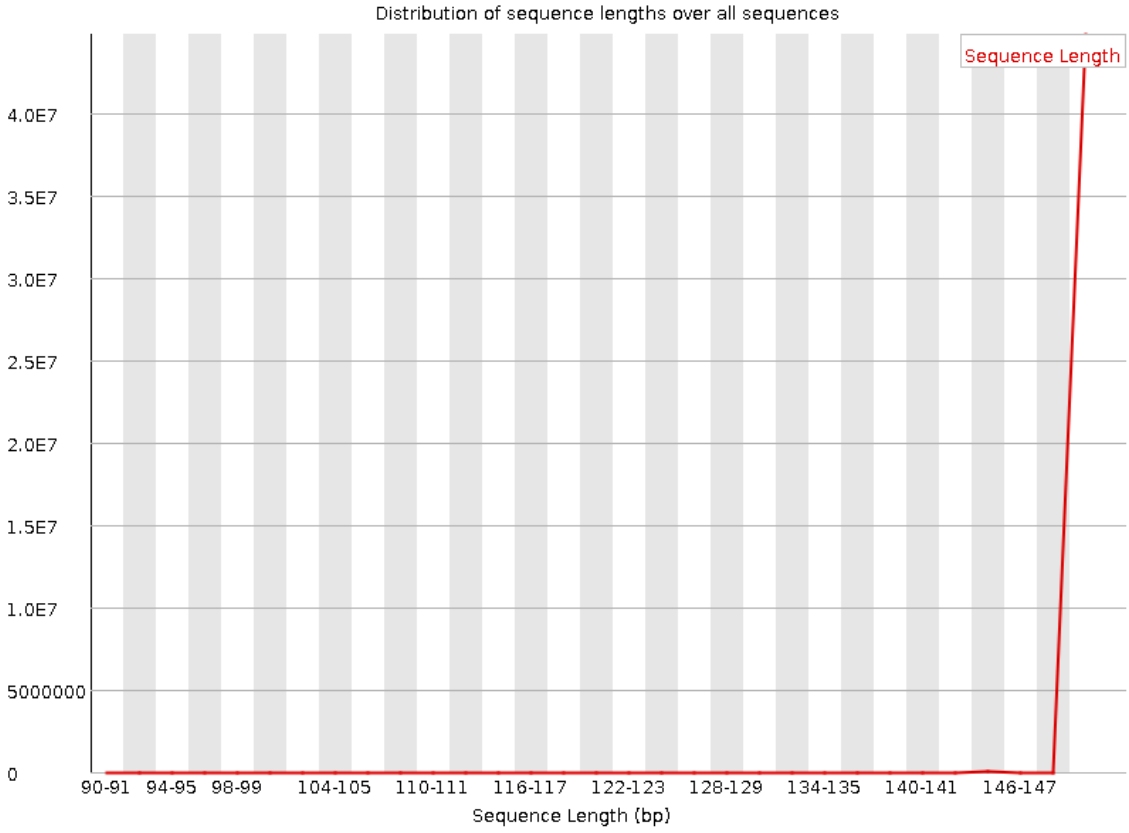


Figura 10. Representação da presença ou ausência de adaptadores.

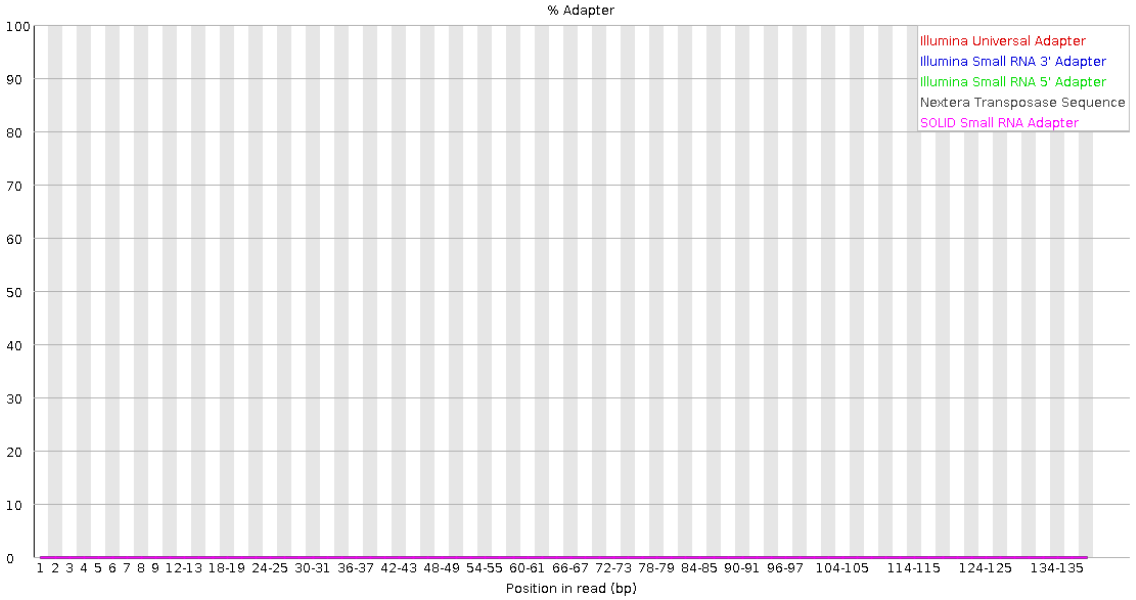


Figura 11. Matriz de identidade gerada através do SDT com sequências do gene L (nt) de todos os representantes da subfamília *Betarhabdovirinae*

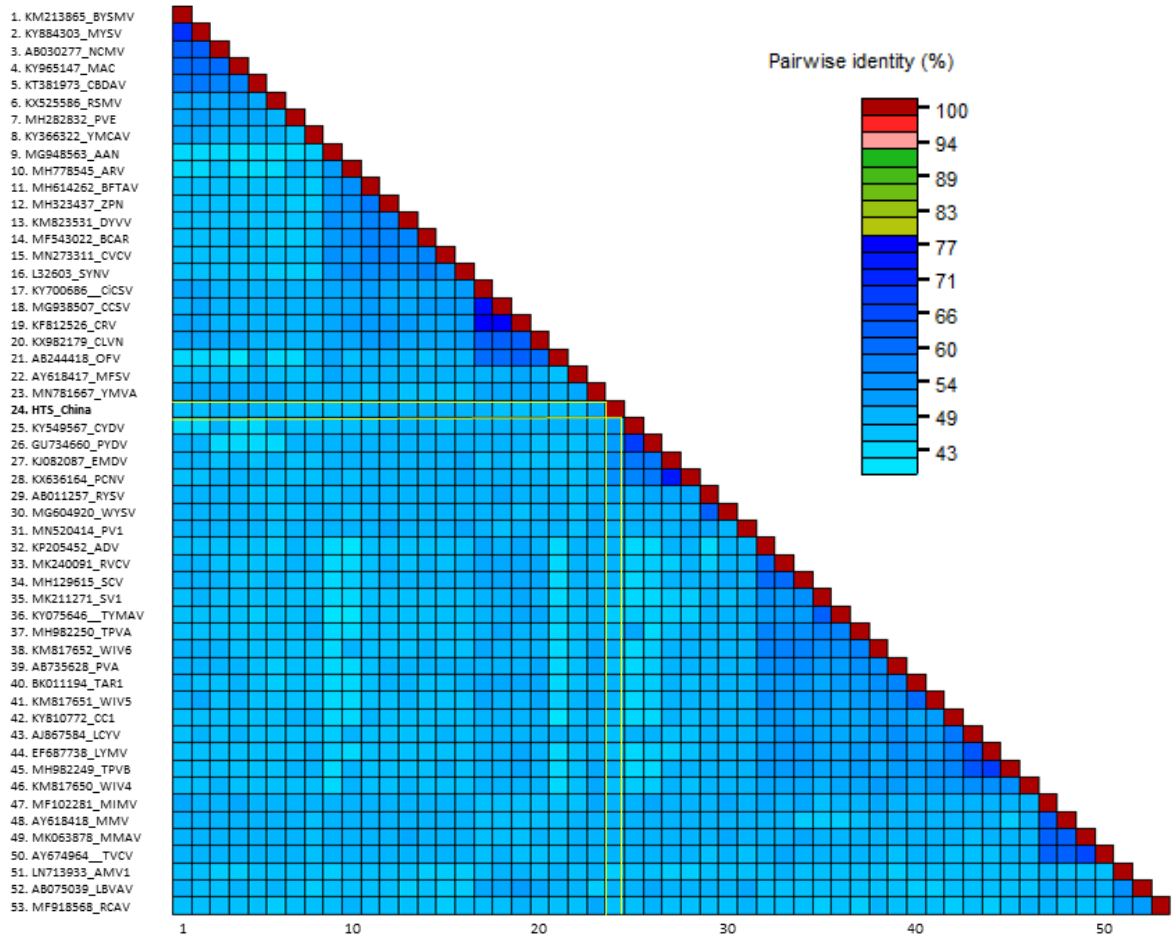


Figura 12. Determinação de blocos conservados dentro da região L (aa) de SRR10480882 e sequências de *Alphanucleorhabdovirus* e *Dichorharvirus*.