



UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE FITOPATOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM FITOPATOLOGIA

VIROMA EM ESPÉCIES ORNAMENTAIS

PEDRO MARETTI BRANT

Brasília – DF
2017

PEDRO MARETTI BRANT

VIROMA EM ESPÉCIES ORNAMENTAIS

Dissertação apresentada à
Universidade de Brasília
como requisito parcial para a
obtenção do título de Mestre
em Fitopatologia pelo
Programa de Pós Graduação
em Fitopatologia

Orientador

Profa. Dra. Rita de Cássia Pereira Carvalho.

Co-orientador

Prof. Dr. Tatsuya Nagata.

BRASÍLIA
DISTRITO FEDERAL - BRASIL
2017

FICHA CATALOGRÁFICA

Brant, Pedro Maretti.

Viroma em Espécies Ornamentais. / Pedro Maretti

Brant. Brasília, 2017.

Número de páginas: 105 p.

Dissertação de mestrado. Programa de Pós-Graduação em Fitopatologia,
Universidade de Brasília, Brasília.

I. NGS - *Nucleorhabdovirus*

II. Universidade de Brasília. PPG/FIT.

III. Viroma em Espécies Ornamentais no Distrito Federal.

À minha esposa.

AGRADECIMENTOS

Ao CNPq e à FAP/DF por financiarem o presente projeto de pesquisa.

À Prof. Dra. Rita de Cássia Pereira Carvalho pela oportunidade e orientação.

A todos do Laboratório de Virologia Animal do CEL-IB por me acolher da melhor forma possível.

Aos Dr. Fernando Lucas Melo e Dra. Rosana Blawid pelo vasto conhecimento compartilhado sempre quando possível.

Ao Prof. Dr. Tatsuya Nagata pela orientação, oportunidade, atenção e, principalmente, extrema confiança.

Trabalho realizado junto ao Departamento de Fitopatologia do Instituto de Ciências Biológicas da Universidade de Brasília, sob orientação da **Professora Doutora Rita de Cássia Pereira Carvalho**, com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

VIROMA DE ESPÉCIES ORNAMENTAIS.

PEDRO MARETTI BRANT

DISSERTAÇÃO APROVADA em ___/___/_____ por:

Profa. Dra. Alice Kazuko Inoue Nagata
Examinador (membro interno)

Dr. Fernando Lucas de Melo
Examinador (membro externo)

Profa. Dra. Rita de Cássia Pereira Carvalho
Orientador (Presidente)

**BRASÍLIA – DISTRITO FEDERAL
BRASIL
2017**

Sumário

LISTA DE TABELAS.....	ix
LISTA DE FIGURAS.....	x
RESUMO GERAL.....	xiii
GENERAL ABSTRACT.....	xv
INTRODUÇÃO GERAL	1
Objetivo Geral.....	4
Objetivos Específicos.....	4
CAPÍTULO 1: Revisão de Literatura.....	5
1.1 Mercado de flores e plantas ornamentais	5
1.2 Vírus em plantas ornamentais	8
1.3 As famílias virais <i>Tombusviridae</i> , <i>Potyviridae</i> e <i>Rhabdoviridae</i>	10
1.4 Sequenciamento de alto desempenho (<i>Next Generation Sequencing</i> – NGS).....	17
1.5 Bioinformática aplicada a virologia vegetal	20
1.6 Referências Bibliográficas.....	24
CAPÍTULO 2: Viroma em espécies ornamentais	28
RESUMO	28
1. INTRODUÇÃO.....	30
2. MATERIAL E MÉTODOS	32
2.1. Coleta e obtenção das amostras de plantas ornamentais a serem analisadas.....	32
2.2. Enriquecimento viral, extração de RNA e envio de um <i>pool</i> de amostras para <i>Next Generation Sequencing</i> - NGS.....	34
2.3. Análise computacional das sequências obtidas via <i>Next Generation Sequencing</i> - NGS	34
2.4. Seleção e montagem do genoma das duas espécies virais: <i>Cariopsis mottle-associated virus</i> (CMaV) e <i>Pachystachys mosaic-associated virus</i> (PMaV) para caracterização molecular	35
2.5. Detecção de seis espécies virais por meio de RT-PCR (<i>Reverse Transcriptase – Polymerase Chain Reaction</i>) nas plantas ornamentais selecionadas	36
2.6. Estratégias para recuperação do genoma completo de <i>Coreopsis mottle-associated virus</i> (CMaV) e <i>Pachystachys mosaic-associated virus</i> (PMaV)	38
2.7. Recuperação das regiões 3' e 5' dos genomas de <i>Cariopsis mottle-associated virus</i> (CMaV) e <i>Pachystachys mosaic-associated virus</i> (PMaV) por meio de RACE (<i>Rapid Amplification of cDNA Ends</i>) ...	41
2.8. Análises Filogenéticas para o <i>Coreopsis mottle-associated virus</i> (CMaV) e <i>Pachystachys mosaic-associated virus</i> (PMaV).....	45
3. RESULTADOS E DISCUSSÃO.....	48
3.1. Plantas ornamentais coletadas no Viveiro I da NOVACAP (Companhia Urbanizadora da Nova Capital do Brasil).....	48
3.2. Sequências obtidas via <i>Next Generation Sequencing</i> - NGS.....	48

3.4. Detecção das sequências virais obtidas via <i>Next Generation Sequencing</i> – NGS no <i>pool</i> e nas amostras de plantas ornamentais individuais.....	58
3.5. Determinação do genoma completo das espécies <i>Coreopsis mottle-associated virus</i> (CMAV) e <i>Pachystachys mosaic-associated virus</i> (PMAV).....	59
3.6. Filogenia das sequências virais <i>Coreopsis mottle-associated virus</i> (CMAV) e <i>Pachystachys mosaic-associated virus</i> (PMAV).....	63
4. CONCLUSÕES GERAIS E PERSPECTIVAS.....	83
5. REFERÊNCIAS BIBLIOGRÁFICAS.....	84

LISTA DE TABELAS

Tabela 1. Plantas ornamentais (nome comum e científico) coletadas para estudo e com base nos sintomas observados no momento da coleta no Viveiro I da NOVACAP (Companhia Urbanizadora da Nova Capital do Brasil).....	32
Tabela 2. <i>Primers</i> usados para detecção de seis espécies virais no <i>pool</i> e nas amostras individuais de plantas ornamentais.	37
Tabela 3. Sequências dos <i>primers</i> utilizados em cada região genômica para as sequências de Coreopsis mottle-associated virus (CMaV) e Pachystachys mosaic-associated virus (PMaV).	40
Tabela 4. <i>Primers</i> utilizados nos protocolos 3' e 5' RACE (<i>Rapid Amplification of cDNA Ends</i>) para determinação de sequências de extremidades das espécies Coreopsis mottle-associated virus (CMaV) e Pachystachys mosaic-associated virus (PMaV).	44
Tabela 5. Identificação das sequências de aminoácidos das espécies virais <i>Sonchus yellow net nucleorhabdovirus</i> (SYNV) e <i>Potato yellow dwarf nucleorhabdovirus</i> (PYDV) utilizados nos alinhamentos <i>pairwise</i> globais.	45
Tabela 6. Sequências de nucleotídeos do genoma completo de vírus da família <i>Rhabdoviridae</i> depositadas no banco de dados (https://www.ncbi.nlm.nih.gov/) e utilizadas nas análises filogenéticas.	46
Tabela 7. Sequências de aminoácidos das proteínas G (glicoproteína), N (nucleocapsídeo) e L (polimerase) de espécies virais da família <i>Rhabdoviridae</i> depositadas no banco de dados (https://www.ncbi.nlm.nih.gov/) e utilizados nas análises filogenéticas.	46
Tabela 8. Número de <i>contigs</i> gerados pelos montadores <i>SPAdes</i> , <i>Velvet</i> (<i>k-mer</i> 29 e <i>k-mer</i> 61) e <i>ABYSS</i> , número total de <i>hits</i> pelo algoritmo <i>tblastx</i> com base no RefSeq viral de setembro de 2016 para cada montador e número de <i>hits</i> específicos com vírus de planta e demais vírus, com base no resultado do algoritmo <i>tblastx</i> usando o RefSeq viral obtido em setembro de 2016.	49
Tabela 9. Quantidade de <i>hits</i> por família de vírus de planta, tamanho de <i>contigs</i> , identidade, cobertura e <i>E-value</i> a partir do algoritmo <i>tblastx</i> com base no RefSeq viral de setembro de 2016. Os <i>contigs</i> que geraram os <i>hits</i> foram produzidos pelos montadores <i>SPAdes</i> , <i>Velvet</i> (<i>k-mer</i> 61 e 29) e <i>ABYSS</i>	50
Tabela 10. Quantidade de <i>contigs</i> , tamanho mínimo e máximo (em pb) e <i>E-value</i> menor e maior para cada sequência viral trabalhada <i>in silico</i> – <i>Sonchus yellow net nucleorhabdovirus</i> (SYNV), <i>Potato yellow dwarf nucleorhabdovirus</i> (PYDV), <i>Alfalfa dwarf cytorhabdovirus</i> (ADV), <i>Jasmine ringspot virus</i> (JRSV), <i>Carrot mottle mimic virus</i> (CMoMV) e <i>Rosa rugosa leaf distortion virus</i> (RrLDV) – em cada montador utilizado (<i>SPAdes</i> , <i>Velvet</i> e <i>ABYSS</i>).	577

LISTA DE FIGURAS

- Figura 1.** Plantas ornamentais coletadas no Viveiro 1 da NOVACAP (Companhia Urbanizadora da Nova Capital do Brasil) e analisadas quanto à presença de vírus por *Next Generation Sequencing* (NGS). As amostras são: A – *Pinanga kuhlii* (palmeira pinanga; com sintoma de mosqueado); B – *Anthurium lindmanianum* (antúrio; com sintoma de amarelecimento e clorose); C – *Impatiens hawkeri* (beijinho; com sintoma de ápice encarquilhado); D – *Pelargonium* sp. (gerânio; com sintoma de anéis cloróticos verdes e amarelos); E – *Epipremnum aureum* (jibóia; com sintoma de amarelecimento); F – *Jasminum nitidum* (jasmim-estrela; com sintoma de anéis cloróticos); G – *Coreopsis lanceolata* (margaridinha; com sintoma de mosqueado); H – *Pachystachys lutea* (camarão-amarelo; com sintoma de mosaico); I – *Neomarica candida* (Lírio-da-praia; com sintoma de amarelecimento); J – *Streptosolen jasmonii* (marianinha; com sintoma de amarelecimento e mosqueado).33
- Figura 2.** Disposição dos *primers* específicos para as sequências: Coreopsis mottle-associated virus (CMaV) e Pachystachys mosaic-associated virus (PMaV) ao longo do respectivo genoma. Os retângulos representam os genes, sendo que os mesmos foram supostos por analogia aos vírus de referência (*Sonchus yellow net virus* (SYNV) e *Potato yellow dwarf virus* (PYDV), respectivamente). As setas representam a localização próxima de cada primer no genoma. Os possíveis genes são referentes às seguintes proteínas: N – nucleocapsídeo; M – proteína de matriz; G – glicoproteína; L – polimerase; sc4 – possível silenciador gênico/proteína de movimento; P – fosfoproteína; X – proteína “X”; Y – possível proteína de movimento. F: *Forward*; R: *Reverse*. Ctg: “*Contig*” equivalente a parte do genoma dividido. GSP: *primers* internos do 5’ RACE (*Rapid Amplification of cDNA Ends*).41
- Figura 3.** Gráficos comparativos entre os montadores (*SPAdes*, *Velvet* e *ABYSS*) mostrando a relação entre vírus de planta e os demais vírus obtidos por meio de sequenciamento NGS (*Next-Generation Sequencing*), utilizando sequências que apresentaram *hits* com sequências virais.....54
- Figura 4.** Gráficos comparativos entre os montadores *SPAdes*, *Velvet* e *ABYSS* referentes a presença de vírus pertencentes a determinadas famílias de vírus de plantas e sua relação com o número total de vírus de planta. Os representantes que não ultrapassaram a margem de 0% não estão apontados nos gráficos.55
- Figura 5.** Eletroforese em gel de agarose (1%) da amplificação de diferentes regiões do genoma de *Coreopsis mottle-associated nucleorhabdovirus* (CMaV), denominadas 1, 2, 3, 4 e 5, a partir de amostras de *Coreopsis lanceolata* (margaridinha). Cinco pares de *primers* (um para cada região de tamanho aproximado de 2.900 pb) foram utilizados para cobrir o genoma de 13.657 pb. M: Marcador Molecular 1 kb (Invitrogen) evidenciado na primeira coluna.60
- Figura 6.** Eletroforese em gel de agarose (1%) da amplificação de diferentes regiões do genoma de *Pachystachys mosaic-associated nucleorhabdovirus* (PMaV), denominados 1, 2, 3 e 4, a partir de amostras de *Pachystachys lutea* (camarão-amarelo). Quatro pares de *primers* (um para cada região de tamanho aproximado de 2.900 pb) foram utilizados para cobrir o genoma de 11.342 pb. M: Marcador Molecular 1 kb (Invitrogen) na primeira coluna. As setas indicam o *amplicon* desejado.....61
- Figura 7.** Esquema com base na sequência obtida via sequenciamento Sanger representando o genoma do *Coreopsis mottle-associated virus* (CMaV) e seus possíveis genes. As setas indicam a posição do início e fim de cada gene.62
- Figura 8.** Esquema com base na sequência obtida via sequenciamento Sanger representando o genoma do *Pachystachys mosaic-associated virus* (PMaV) e seus possíveis genes. As setas indicam a posição do início e fim de cada gene.....62
- Figura 9.** Matriz de porcentagem de identidade entre a sequência do genoma completo de *Coreopsis mottle-associated virus* (CMaV) obtida via sequenciamento Sanger de amostras de *Coreopsis lanceolata* com sequências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de nucleotídeo.66
- Figura 10.** Matriz de porcentagem de identidade entre as sequências de diferentes rhabdovírus e o *Coreopsis mottle-associated virus* (CMaV) a nível da proteína L (polimerase) obtida via sequenciamento Sanger de amostras de *Coreopsis lanceolata* com sequências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de aminoácido.....67

Figura 11. Matriz de porcentagem de identidade entre as sequências de diferentes rhabdovírus e o Coreopsis mottle-associated virus (CMaV) a nível da proteína N (nucleocapsídeo) obtida via sequenciamento Sanger de amostras de <i>Coreopsis lanceolata</i> com sequências de diferentes rhabdovírus obtidas diretamente do <i>National Center for Biotechnology Information</i> (NCBI) a nível de aminoácido.....	68
Figura 12. Matriz de porcentagem de identidade entre as sequências de diferentes rhabdovírus e o Coreopsis mottle-associated virus (CMaV) a nível da proteína G (glicoproteína) obtida via sequenciamento Sanger de amostras de <i>Coreopsis lanceolata</i> com sequências de diferentes rhabdovírus obtidas diretamente do <i>National Center for Biotechnology Information</i> (NCBI) a nível de aminoácido.....	69
Figura 13. Árvore filogenética com base no algoritmo <i>Maximum Likelihood</i> (1000 repetições <i>bootstrap</i>) do genoma completo do vírus Coreopsis mottle-associated virus (CMaV) obtida via sequenciamento Sanger de amostras de <i>Coreopsis lanceolata</i> com sequências de diferentes rhabdovírus obtidas diretamente do <i>National Center for Biotechnology Information</i> (NCBI) a nível de nucleotídeo. A espécie viral CMaV encontra-se no mesmo clado de <i>Sonchus yellow net nucleorhabdovirus</i> e fazendo parte do gênero <i>nucleorhabdovirus</i>	70
Figura 14. Árvore filogenética com base no algoritmo <i>Maximum Likelihood</i> (1000 repetições <i>bootstrap</i>) da proteína L (polimerase) do vírus Coreopsis mottle-associated virus (CMaV) obtida via sequenciamento Sanger de amostras de <i>Coreopsis lanceolata</i> com sequências de diferentes rhabdovírus obtidas diretamente do <i>National Center for Biotechnology Information</i> (NCBI) a nível de aminoácido. A espécie viral CMaV encontra-se no mesmo clado do <i>Sonchus yellow net nucleorhabdovirus</i> , indicando que CMaV faça parte do gênero <i>Nucleorhabdovirus</i>	71
Figura 15. Árvore filogenética com base no algoritmo <i>Maximum Likelihood</i> (1000 repetições <i>bootstrap</i>) da proteína G (glicoproteína) do vírus Coreopsis mottle-associated virus (CMaV) obtida via sequenciamento Sanger de amostras de <i>Coreopsis lanceolata</i> com sequências de diferentes rhabdovírus obtidas diretamente do <i>National Center for Biotechnology Information</i> (NCBI) a nível de aminoácido. A espécie viral CMaV encontra-se no mesmo clado do <i>Sonchus yellow net nucleorhabdovirus</i> , indicando que CMaV faça parte do gênero <i>Nucleorhabdovirus</i>	72
Figura 16. Árvore filogenética com base no algoritmo <i>Maximum Likelihood</i> (1000 repetições <i>bootstrap</i>) da proteína N (nucleocapsídeo) do vírus Coreopsis mottle-associated virus (CMaV) obtida via sequenciamento Sanger de amostras de <i>Coreopsis lanceolata</i> com sequências de diferentes rhabdovírus obtidas diretamente do <i>National Center for Biotechnology Information</i> (NCBI) a nível de aminoácido. A espécie viral CMaV encontra-se no mesmo clado do <i>Sonchus yellow net nucleorhabdovirus</i> , indicando que CMaV faça parte do gênero <i>Nucleorhabdovirus</i>	73
Figura 17. Matriz de porcentagem de identidade entre a sequência do genoma completo de Pachystachys mosaic-associated virus (PMaV) obtida via sequenciamento Sanger de amostras de <i>Pachystachys lutea</i> com sequências de diferentes rhabdovírus obtidas diretamente do <i>National Center for Biotechnology Information</i> (NCBI) a nível de nucleotídeo.	75
Figura 18. Matriz de porcentagem de identidade entre as sequências de diferentes rhabdovírus e o Pachystachys mosaic-associated virus (PMaV) a nível da proteína L (polimerase) obtida via sequenciamento Sanger de amostras de <i>Pachystachys lutea</i> com sequências de diferentes rhabdovírus obtidas diretamente do <i>National Center for Biotechnology Information</i> (NCBI) a nível de aminoácido.....	76
Figura 19. Matriz de porcentagem de identidade entre as sequências de diferentes rhabdovírus e o Pachystachys mosaic-associated virus (PMaV) a nível da proteína N (nucleocapsídeo) obtida via sequenciamento Sanger de amostras de <i>Pachystachys lutea</i> com sequências de diferentes rhabdovírus obtidas diretamente do <i>National Center for Biotechnology Information</i> (NCBI) a nível de aminoácido. ..	77
Figura 20. Matriz de porcentagem de identidade entre as sequências de diferentes rhabdovírus e o Pachystachys mosaic-associated virus (PMaV) a nível da proteína G (glicoproteína) obtida via sequenciamento Sanger de amostras de <i>Pachystachys lutea</i> com sequências de diferentes rhabdovírus obtidas diretamente do <i>National Center for Biotechnology Information</i> (NCBI) a nível de aminoácido. ..	78
Figura 21. Árvore filogenética com base no algoritmo <i>Maximum Likelihood</i> (1000 repetições <i>bootstrap</i>) do genoma completo do vírus Pachystachys mosaic-associated virus (PMaV) obtida via sequenciamento Sanger de amostras de <i>Pachystachys lutea</i> com sequências de diferentes rhabdovírus obtidas diretamente do <i>National Center for Biotechnology Information</i> (NCBI) a nível de nucleotídeo. A espécie viral PMaV encontra-se no mesmo clado de <i>Potato yellow dwarf nucleorhabdovirus</i> (PYDV) e fazendo parte do gênero <i>Nucleorhabdovirus</i>	79

Figura 22. Árvore filogenética com base no algoritmo *Maximum Likelihood* (1000 repetições *bootstrap*) da proteína L (polimerase) do vírus *Pachystachys* mosaic-associated virus (PMAV) obtida via sequenciamento Sanger de amostras de *Pachystachys lutea* com sequências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de aminoácido. A espécie viral PMAV encontra-se no mesmo clado de *Potato yellow dwarf nucleorhabdovirus* (PYDV) e fazendo parte do gênero *Nucleorhabdovirus*.80

Figura 23. Árvore filogenética com base no algoritmo *Maximum Likelihood* (1000 repetições *bootstrap*) da proteína N (nucleocapsídeo) do vírus *Pachystachys* mosaic-associated virus (PMAV) obtida via sequenciamento Sanger de amostras de *Pachystachys lutea* com sequências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de aminoácido. A espécie viral PMAV encontra-se no mesmo clado de *Potato yellow dwarf nucleorhabdovirus* (PYDV) e fazendo parte do gênero *Nucleorhabdovirus*.81

Figura 24. Árvore filogenética com base no algoritmo *Maximum Likelihood* (1000 repetições *bootstrap*) da proteína G (glicoproteína) do vírus *Pachystachys* mosaic-associated virus (PMAV) obtida via sequenciamento Sanger de amostras de *Pachystachys lutea* com sequências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de aminoácido. A espécie viral PMAV encontra-se no mesmo clado de *Potato yellow dwarf nucleorhabdovirus* (PYDV) e fazendo parte do gênero *Nucleorhabdovirus*.82

RESUMO GERAL

BRANT, Pedro Maretti. **Viroma em Espécies Ornamentais**. 2017. 105 p. Dissertação (Mestrado em Fitopatologia) – Universidade de Brasília, Brasília, DF.

A floricultura no Brasil teve início como atividade secundária da fruticultura nos estados de São Paulo e Santa Catarina no início do século XX e atualmente apresenta um crescimento anual de exportação maior que a média mundial. O Distrito Federal (DF) é o décimo em valor de mercado no país, sendo um mercado altamente promissor por apresentar a maior renda *per capita* do país. Com a expansão do setor, aumento na incidência das doenças e a constante introdução de mudas e sementes exóticas no país, fatores de cultivo, manejo e sanidade vegetal podem constituir um entrave à produção de ornamentais no país e DF. Ainda assim, pesquisas direcionadas a detecção precisa de espécies virais ocorrendo em plantas ornamentais são incipientes no país e escassas no DF. O estudo de viroma em combinação com as estratégias de *Next Generation Sequencing* (NGS), tendo o apoio das metodologias de bioinformática, têm possibilitado uma eficiente descrição de diferentes organismos presentes em diferentes amostras biológicas nos mais diversificados ambientes e proporcionado avanços significativos em diferentes campos das ciências, incluindo a virologia vegetal. Neste trabalho, um *pool* composto por dez espécies de plantas ornamentais provenientes do Viveiro I da NOVACAP (Companhia Urbanizadora da Nova Capital do Brasil) foi submetido ao sequenciamento NGS. Após análise dos dados gerados, por meio de diferentes técnicas de bioinformática, sequências novas foram encontradas, indicando seis prováveis novas espécies virais em diferentes gêneros e famílias. Duas prováveis espécies novas de *Nucleorhabdovirus* (família *Rhabdoviridae*) foram identificadas, sendo uma delas em *Pachystachys lutea* (camarão-amarelo) e a outra em *Coreopsis lanceolata* (margaridinha). Nesta mesma ornamental, margaridinha, identificou-se também uma provável espécie nova do gênero *Umbravirus* (família *Tombusviridae*). Outro representante da família *Tombusviridae*, neste caso uma espécie classificada no gênero *Pelarspovirus*, foi identificada em *Jasminum nitidum* (jasmim estrela). As demais espécies, uma do gênero *Cytorhabdovirus* e outra uma do gênero *Potyvirus*, foram detectadas no *pool* de amostras. Dentre essas prováveis espécies novas de vírus, duas sequências detectadas nas plantas ornamentais *C. lanceolata* e *P. lutea* apresentaram maior proximidade com os vírus *Sonchus yellow net nucleorhabdovirus* (SYNV) e *Potato yellow dwarf nucleorhabdovirus* (PYDV), respectivamente, e foram denominadas como *Coreopsis mottle-associated virus* (CMaV) e

Pachystachys mosaic-associated virus (PMaV). Para CMaV e PMaV, o genoma total foi recuperado mediante amplificação genômica nas amostras originais utilizando *primers* internos sobrepostos, sendo as extremidades 3' e 5' recuperadas através da técnica RACE (*Rapid Amplification of cDNA Ends* - RACE). Após estes procedimentos, sequências foram obtidas via sequenciamento Sanger, confirmando a alta identidade com a sequência montada *in silico* pelo NGS. Com a sequência completa dos dois vírus procedeu-se análises filogenéticas. A partir de análises comparativas via alinhamento múltiplo de sequências (MSA), alinhamento *pairwise* e filogenéticas, propôs-se que CMaV e PMaV sejam prováveis espécies novas para o gênero *Nucleorhabdovirus*, tendo em vista a identidade nucleotídica de 56,99% entre os genomas de CMaV e SYNV e 46,99% entre PMaV e PYDV.

Palavras-chave: Bioinformática, Filogenia, NGS, *Nucleorhabdovirus*, Ornamentais, *Rhabdoviridae*, Viroma.

Orientador – Rita de Cássia Pereira Carvalho - Universidade de Brasília; Co-orientador – Tatsuya Nagata – Universidade de Brasília.

GENERAL ABSTRACT

BRANT, Pedro Maretti. **Virome in Ornamental Plants**. 2017. 105 p. Dissertation (Master in Plant Pathology – University of Brasilia, Brasília, DF, Brazil)

In Brazil, the floriculture agribusiness started as a minor activity in the São Paulo and Santa Catarina states in the early 20th century and currently its annual market growth is bigger than worldwide's mean. The Federal District (DF) is ranked in the 10th place in terms of market value in Brazil, being a highly promising market due to the high average per capita income of its population. With the expansion of this sector, it was observed a significant increase in the incidence of diseases due to the constant transit and introduction into the country of exotic plant material and seeds. In addition, the implementation of new crop management practices is also having direct effects on disease outbreaks. These factors may affect the sustainability of the ornamental plant agribusiness sector in the country. Research efforts aiming to develop accurate diagnostic tools for viral species occurring in ornamental plants are yet incipient in the country. The virome study employing *Next Generation Sequencing* (NGS) techniques, combined with bioinformatics methods, has been considered as the most powerful strategy aiming to exploit the genetic diversity of organisms present in any biological sample and allowed significant advances in different fields of science, including plant virology. For this work a *pool* of ten ornamental plants from NOVACAP (*Companhia Urbanizadora da Nova Capital do Brasil*) Nursery 1 was sent to NGS. Posterior to data analysis, the sequences obtained were compared to reference genomes from DNA data banks and assembled, producing potential six plant virus sequences in different genera and families. Two probable new species from *Nucleorhabdovirus* (*Rhabdoviridae* family) were identified, one in *Pachystachys lutea* and other in *Coreopsis lanceolata*. In this ornamental plant, *C. lanceolata*, it was identified as well a probable new species of the genus *Umbravirus* (*Tombusviridae* family). Other member of the *Tombusviridae* family, now from the *Pelarspovirus* genus, was identified in *Jasminum nitidum*. Two of the assembled species, one of the *Cytorhabdovirus* genus and other from the *Potyvirus* genus, were detected in the *pool* only. Among these probable new viral species, two sequences that were detected in the ornamental plants *C. lanceolata* and *P. lutea* were close to the *Sonchus yellow net nucleorhabdovirus* (SYNV) and *Potato yellow dwarf nucleorhabdovirus* (PYDV) viruses, respectively, and were named as *Coreopsis mottle-associated virus* (CMaV) and *Pachystachys mosaic-associated virus* (PMaV). For CMaV and PMaV, the total genome was

recovered by amplifying the genome from the original samples using sintesized *primers*, while the 3' and 5' ends with the RACE (Rapid Amplification of cDNA Ends) technique. Posterior to these assays, sequences were obtained from Sanger sequencing method, confirming the high identity with the *in silico* assembled sequence from NGS. With the complete sequence of both viruses, phylogenetics assays were made. With the multiple sequence alignment (MSA), *pairwise* alignment and phylogenetics, we propose that CMaV and PMaV are probable new viruses from the *Nucleorhabdovirus* genus, as the nucleotide identity of the genome were 56.99% from CMaV and SYNIV and 46.99% from PMaV and PYDV.

Keywords: Bioinformatics, Ornamental plants, NGS, *Nucleorhabdovirus*, Phylogeny, *Rhabdoviridae*, Virome.

Advisory board: Rita de Cássia Pereira Carvalho – Universidade de Brasília (Advisor), Tatsuya Nagata – Universidade de Brasília (Co-advisor).

INTRODUÇÃO GERAL

A comercialização de plantas ornamentais vem ganhando espaço no Brasil, com o desenvolvimento do mercado internacional e o conseqüente aumento de importações e exportações. Como as plantas ornamentais são altamente versáteis em sua funcionalidade e uso, o mercado cresce em ritmo acelerado. Internacionalmente, a exportação atingiu o patamar de U\$ 60 bilhões, sendo liderada pela Holanda. Desde o início do século XXI, alguns países em desenvolvimento entraram no *ranking* de maiores exportadores de ornamentais, incluindo o Brasil (Ruud & Nico, 2005).

O mercado de flores e plantas ornamentais no Brasil teve início como atividade secundária por imigrantes. Somente após metade do século XX, a indústria passou a ter força e alguma visibilidade no mercado externo. Desde então, somam-se crescimentos em níveis significativos, movimentando quase 6 bilhões de reais em 2014 (Junqueira & Peetz, 2015). O Distrito Federal é o terceiro maior consumidor do país, sendo abastecido por produtores externos, como São Paulo, e pela produção local, que já chega a cobrir 15% da demanda. Além disso, a produção no DF ocupa a décima posição no país e está em ascensão, gerando empregos e favorecendo o comércio local (Alonso & Silva, 2009).

Com o aumento da produção de plantas ornamentais no Brasil e no mundo, crescem os números de casos de doenças afetando a produtividade. O uso de material propagativo infectado indiscriminadamente é uma potencial causa para a disseminação de doenças e patógenos em cultivos de plantas ornamentais (Alexandre, Duarte et al. 2010). A associação de vírus e plantas ornamentais é retratada em um dos primeiros relatos históricos da Virologia, a quebra de coloração das flores de tulipas na Holanda por um vírus, no século XVII, gerando o fenômeno conhecido como “Tulipomania” (Hull, 2013).

Com o desenvolvimento e amadurecimento do mercado de flores e plantas ornamentais no Brasil e mundo, a necessidade em garantir a alta produção promoveu uma série de pesquisas envolvendo a identificação e caracterização de patógenos em ornamentais. Em relação a vírus, algumas importantes espécies para diferentes grandes culturas já foram detectadas em ornamentais, como *Bean yellow mosaic virus* – BYMV (família *Potyviridae*), *Cucumber mosaic virus* – CMV (família *Bromoviridae*) e *Groundnut ringspot virus* – GRSV (família *Bunyaviridae*) infectando bulbos importados, indicando que plantas ornamentais podem servir como hospedeiras e possíveis fonte de inóculo. Isso gera preocupação para as

questões de importação e exportação, dando, assim, enfoque para a área de fitossanidade nesse setor (Alexandre et al., 2010).

Vírus são parasitas intracelulares obrigatórios e, portanto, necessitam de um hospedeiro para completar o seu ciclo. No caso de vírus de plantas, muitas são as espécies conhecidas por acometer e promover doenças graves em plantas de interesse econômico, como é o caso de espécies do gênero *Potyvirus* (Hull, 2013). Para algumas espécies de outros gêneros virais, no entanto, os estudos são ainda escassos, devido ao fato destas espécies não promoverem (ou por se desconhecer), até então, doenças de grande impacto ou por não fazerem parte de famílias virais que abrigam espécies economicamente importantes e amplamente estudadas, como *Geminiviridae*, *Potyviridae*, *Bunyaviridae* e *Closteroviridae*. No caso dos vírus de planta da família *Rhabdoviridae*, por exemplo, existem alguns estudos no país, entretanto há poucos genomas completos sequenciados e disponíveis em bancos de dados, dificultando a classificação dos vírus neste grupo (Jackson et al., 2005). O mesmo é observado para a família *Tombusviridae*, onde relatos de espécies virais acometendo plantas, como ornamentais e cenoura (*Daucus carota*) já foram feitos (Mollov et al., 2013), entretanto poucos genomas completos estão depositados em banco de dados. Alguns estudos relatam a presença de vírus em plantas ornamentais, entretanto a caracterização das espécies, e até mesmo classificação em nível de gênero, ainda são escassos e estudos moleculares são raros ou ausentes no DF e no país (Alexandre, Duarte et al. 2010).

De acordo com Mokili et al. (2012), os vírus são considerados a entidade mais abundante na biosfera, entretanto menos de 1% da diversidade viral é conhecida. Comparando esta proporção com dados fornecidos pelo *International Committee on Taxonomy of Viruses* (ICTV) que descreve 3.186 espécies virais (ICTV, 2016), percebe-se que pouco se conhece sobre a diversidade viral. Algumas explicações para tal fato advém de características peculiares dos vírus que dificultam a sua caracterização, por meio de técnicas clássicas, como pouca diversidade morfológica, tamanho diminuto e ausência de sinais na planta. Com o advento do método da metagenômica, espera-se que esse número de espécies possa crescer significativamente nos próximos anos (Mokili et al., 2012).

Novas tecnologias podem revolucionar qualquer área científica. Assim foi na virologia com a microscopia eletrônica de transmissão, a *Polimerase Chain Reaction* (PCR) e, agora, com o *High-Throughput Sequencing*, ou *Next Generation Sequencing* (NGS). Estas metodologias inovadoras mudam a forma de como os estudos relacionados a vírus são realizados, particularmente em áreas como sequenciamento, evolução, ecologia, transcriptômica e metagenômica (Mokili, Rohwer et al. 2012). Nessa última, o NGS tem

contribuído para a ascensão na descoberta de novos vírus, mesmo, em muitos casos, não sendo possível, em um primeiro momento, cumprir os Postulados de Koch e correlacioná-los a doenças específicas (Mokili et al., 2012, Radford et al., 2012).

Com o advento da tecnologia, obter e gerar centenas de milhares de dados é, hoje, considerado uma tarefa relativamente fácil. Os ensaios posteriores, notadamente os de bioinformática, são criticamente importantes no processo de análise de dados e geração de resultados concisos. Ainda assim, obter e analisar os dados tornou-se um processo automatizado, o que compromete a qualidade na interpretação destes. Estudar, organizar e interpretar tamanho número de dados gerados carece de um estudo prévio aprofundado e uma noção em técnicas e métodos em bioinformática (Radford et al., 2012).

A quantidade de dados gerada a cada sequenciamento via NGS está, cada vez mais, maior e mais complexa, gerando um verdadeiro “gargalo” em relação aos resultados analisados e discutidos. A incapacidade de lidar com tamanho número de informações é uma problemática atual, para a qual há, portanto, a necessidade de pesquisas para novas metodologias e softwares que lidem com o tamanho e a dificuldade dos resultados obtidos via sequenciamento (Scholz et al., 2012).

A tecnologia do sequenciamento NGS permite a detecção, identificação e descoberta de novos vírus em plantas de uma maneira direta e sem conhecimento prévio do agente ou da doença por ele causada. Assim, torna-se possível ter noções de populações virais em um determinado ecossistema ou grupo de plantas selecionadas. A tecnologia tem a vantagem de atuar em casos onde não há sequências de referência, porém a falta de uma cobertura significativa prévia pode acarretar em possíveis erros ao longo da interpretação de dados (Prabha et al., 2013). A determinação da diversidade de vírus em plantas ornamentais, sua população e seus hospedeiros pode colaborar com o banco de dados referentes a esse tipo de cultura, colaborando, assim, com futuros trabalhos envolvendo metagenômica e NGS.

HIPÓTESE/JUSTIFICATIVA DO TRABALHO

Existem espécies virais conhecidas para a ciência e/ou prováveis espécies novas associadas a plantas ornamentais cultivadas no Distrito Federal.

Objetivo Geral

Determinar a diversidade viral em dez espécies ornamentais provenientes do Viveiro I da Companhia Urbanizadora da Nova Capital do Brasil (NOVACAP) no Distrito Federal por meio de metagenômica.

Objetivos Específicos

- Identificar espécies virais ocorrendo em amostras de dez espécies de plantas ornamentais provenientes do Viveiro I da NOVACAP;
- Detectar as possíveis espécies virais encontradas em plantas ornamentais específicas;
- Testar e avaliar três diferentes montadores de sequências: *SPAdes*, *Velvet* e *ABYSS*;
- Realizar análises filogenéticas e de comparação de sequências destes vírus ocorrendo em espécies ornamentais;
- Recuperar o genoma completo de prováveis espécies novas e contribuir com informações de sequências virais para banco de dados e para Taxonomia de vírus de plantas.

CAPÍTULO 1

REVISÃO DE LITERATURA

1. Mercado de flores e plantas ornamentais

Plantas consideradas ornamentais são aquelas com algum apelo estético para um determinado setor comercial. São cultivadas e/ou comercializadas com ênfase em seu impacto visual, mas também podem apresentar funcionalidade. Entre vários exemplos de funcionalidade que podem ser citados, encontra-se algumas espécies arbustivas. Amplamente utilizadas em projetos paisagísticos como cercas-vivas, espécies arbustivas propiciam privacidade as pessoas e conforto ao ambiente, além do efeito decorativo (Paiva 2008).

De um modo geral, plantas ornamentais são conhecidas pela beleza de suas flores e/ou folhas. Estas plantas são amplamente produzidas em diferentes países do mundo destinando-se a arquitetura de interiores e paisagismo. Existem indícios que desde os primórdios da humanidade espécies de plantas já eram cultivadas com finalidade ornamental. Um destes exemplos é o lírio-branco (*Lilium candidum*), cujo registro mais antigo refere-se a pinturas da civilização Minoica (na ilha de Creta) considerada a mais antiga civilização que se tem registro na Europa (2700-1420 a.C) (Paiva, 2008).

Muitas espécies de plantas ornamentais podem ser usadas em vasos individuais ou para flores de corte, e ainda em canteiros na composição de jardins, que retratam não apenas a beleza de suas cores, flores e folhas, mas acompanham os fatos históricos de uma determinada época ao longo da história da humanidade, ilustrando a importância destas espécies para o homem moderno. Com a descoberta da América em 1492 pelos espanhóis, novas espécies de ornamentais foram introduzidas no Velho Mundo e as espécies tropicais puderam ser conhecidas e a cada dia vem ganhando espaço no mercado (Paiva, 2008).

A atratividade das flores, aliada a forma e colorido das folhas e, em alguns casos, características peculiares da planta, conduziram o homem no processo seletivo de aprimoramento de qualidades desejáveis nestas espécies, desde a coloração, tamanho e resistência a pragas, patógenos e condições adversas. Assim, muito do que se cultiva atualmente, guarda apenas resquícios do seu cultivo e estado original (Paiva, 2008).

Para o comércio internacional, o mercado de plantas ornamentais é dividido em três grandes setores: flores e folhas de corte, plantas envasadas e plantas para jardim. Atualmente, de acordo com o *International Trade Statistics Database* (UM Comtrade, 2014), a indústria de plantas ornamentais encontra-se em significativo crescimento, apresentando um valor mundial de U\$ 60 bilhões em 2003, com um crescimento anual estimado em 9% desde então. O continente que mais exporta é a Europa, liderada pela Holanda com 52% do mercado de flores de corte em 2013. Em seguida, encontra-se a América do Sul representada pela Colômbia (15%) e Equador (9%). A partir de 2003, um aumento de participação de mercado de alguns países, como Quênia e Etiópia, com 7% e 2% em 2013, respectivamente vem sendo observado (UNComtrade, 2014). Tem-se, desde o início do século 21, portanto, um aumento na competitividade internacional, dando importância a alguns países que, antes, eram restritos a importação e/ou comércio local. Envolvendo todos os três setores do mercado de ornamentais (flores, plantas envasadas e plantas para jardim), o Brasil compreende 6% das exportações (Ruud & Nico, 2005) No Brasil, um dos aspectos positivos ao cultivo de ornamentais consiste nas condições climáticas que favorecem o cultivo tanto de espécies de clima temperado quanto de clima tropical (Junqueira & Peetz, 2015).

O mercado brasileiro de plantas ornamentais surgiu no final do século XIX como atividade secundária da fruticultura nos estados de São Paulo e Santa Catarina. A indústria nacional, no entanto, apresentou-se como potência comercial significativa somente a partir da década de 1950, com ênfase na fundação da Cooperativa Agropecuária de Holambra, em 1972, onde permanece como principal foco de produção no país (Junqueira & Peetz, 2008). Atualmente, o mercado de plantas ornamentais no país apresenta um crescimento de 15%, ou seja, consideravelmente acima da média internacional, de 9%. Em 2014, o mercado de plantas ornamentais conseguiu atingir o patamar de R\$ 5,7 bilhões, sendo o valor e o crescimento tipicamente associados ao aumento do poder aquisitivo da população, melhoria na infraestrutura e maior participação no comércio internacional. Além disso, o mercado é abastecido por diferentes comércios que garantem movimentação ao longo de todo o ano: floricultura, paisagismo, datas comemorativas, decoração e funerária (ABAFEP, 2014).

Dados da IBRAFLOR (Instituto Brasileiro de Floricultura) (2014) destacam a capacidade de geração de empregos (8 funcionários por hectare - ha) e renda na floricultura. O Brasil movimentou R\$ 5,2 bilhões em 2013 e R\$ 5,7 bilhões em 2014 em valor de venda total, e o consumo interno, embora ainda pequeno (US\$ 4,7 por habitante/ano) tem potencial para, no mínimo, atingir o dobro do atual nos próximos anos.

Segundo dados do Instituto Brasileiro de Floricultura (IBRAFLO, 2014), a produção de plantas ornamentais no Brasil foi realizada em 2014 em uma área total de 14.163 hectares (ha), com a maioria cultivada ao ar livre e destinada a plantas de jardim (árvores, arbustos e forrações). A maior área de produção é encontrada na região Sudeste, com 8.561 ha, seguido das regiões Sul, Nordeste e Norte, com 2.714, 2.027 e 861 ha, respectivamente. A região Centro-Oeste, apesar de apresentar uma área menor (829 ha), gerou mais empregos e possui um maior número de pontos de varejo em relação à região Norte. A região, também, é a terceira na produção de plantas para jardim, com 9,95% do mercado nacional. As regiões Sudeste e Sul encabeçam a lista, com 60,41 e 16,57%, respectivamente. Em relação às flores de corte e flores envasadas, o mercado é dominado pela região Sudeste, ambos os setores com fatias acima de 80% (Junqueira & Peetz, 2015).

A região Centro-Oeste é representada por Goiás (GO) e Distrito Federal (DF), que estão posicionados no 9º e 10º lugares em valor de mercado, respectivamente. O DF possui 196 produtores e 360 pontos de varejo, responsáveis por gerar em 2014, um total de R\$ 125.072.000 de valor de mercado (Junqueira & Peetz, 2015, IBRAFLO, 2014). A atividade de plantas ornamentais é relativamente recente no DF, apresentando um impulso com a criação do polo denominado Central Flores (Associação Brasileira dos Produtores de Flores e Plantas), localizado e associado ao CEASA-DF, e atualmente em franca expansão. A Central Flores foi criada em 2001 e conta com vários produtores do Distrito Federal e Entorno, além de outros produtores/empresas não associados que atualmente são responsáveis por atender uma fração da demanda de flores no Distrito Federal que ocupa a posição de 3º maior consumidor nacional e o 1º consumidor per capita do país, sendo a 2º cidade em eventos atrás apenas de São Paulo. Estima-se que 15% da demanda local seja abastecida pelos produtores locais, com o restante obtido pelo polo de São Paulo. Ainda assim, é tratado como um mercado altamente promissor por apresentar uma população com a maior renda per capita do Brasil (Alonso & Silva, 2009). No DF, outro importante polo produtor e distribuidor é a empresa estatal Companhia Urbanizadora da Nova Capital do Brasil (NOVACAP). O Viveiro I da NOVACAP é responsável pelo abastecimento, produção e implantação de todas as plantas ornamentais em canteiros públicos e residências oficiais no DF (NOVACAP, 2017).

Apesar do crescimento anual elevado e promissor, o mercado nacional tem o seu desenvolvimento atrasado por alguns entraves, como infraestrutura ainda precária em alguns setores e regiões, tributação exagerada, escassez de mão de obra e crise no abastecimento de água no Sudeste (IBRAFLO, 2014). Ainda, com a expansão do setor, viu-se um aumento significativo na incidência de doenças em plantas ornamentais cultivadas no país. A contínua

importação de material de diferentes regiões do mundo agrava a situação, tornando o exame fitossanitário de plantas ornamentais essencial para o controle de entrada de patógenos e pragas no país. Uma das causas atribuídas para a entrada do patógeno ocorre pela entrada indiscriminada de material propagativo infectado entre regiões do país (Alexandre et al., 2010).

2. Vírus em plantas ornamentais

Estudos de doenças em plantas ornamentais nos Estados Unidos da América (EUA) foram iniciados por volta de 1880. Até 1940, no entanto, plantas ornamentais não eram reconhecidas como atividade econômica de importância. A partir desta década, então, a área de pesquisa de doenças em plantas ornamentais vem crescendo e adquirindo novos espaços na Fitopatologia. Devido a indústria de flores e plantas ornamentais ser consideravelmente distinta das demais grandes culturas (ciclo de vida, propagação e comércio, entre outros), pesquisas envolvendo doenças também devem ser realizadas, já que a relação planta-patógeno tende a apresentar características únicas para cada espécie de planta e patógeno. Uma dessas características é problemática com a exportação/importação de material infectado por vírus latentes (Baker & Linderman, 1979).

Um dos primeiros relatos de doença causada por vírus na história da Fitopatologia ocorreu em uma espécie ornamental classificada na família Liliaceae, atualmente bastante difundida: a tulipa. Na primeira metade do século XVII, o fenômeno conhecido como “Tulipomania” impulsionou uma das primeiras “bolhas” econômicas já reportadas. Recém introduzida na Europa, a tulipa passou a ser bastante apreciada e vendida por preços exorbitantes. Muito do sucesso da planta ornamental foi devido a uma infecção viral causada por uma espécie de *Potyvirus*, *Tulip breaking virus* (TBV), afetando os bulbos de tulipas ocasionando variegação de coloração nas pétalas. Esse sintoma contribuiu para que o produto final fosse amplamente apreciado pela população (Hull, 2013). Mais de 200 anos depois, a Virologia se consolidou como Ciência e várias espécies virais foram descritas em espécies vegetais, principalmente em grandes culturas, hortaliças e fruteiras, entretanto para plantas ornamentais, embora já se conheça algumas espécies, acredita-se que exista uma diversidade viral ainda a ser conhecida (Alexandre, Duarte *et al*, 2010).

Com o aumento da importância econômica da produção de flores e plantas ornamentais no Brasil e no mundo, estudos referentes a vírus presentes em espécies ornamentais vêm surgindo com o intuito de catalogar e gerar novas medidas de controle, como exclusão, para garantir o aumento contínuo da produção. Especificamente no Brasil, alguns estudos

envolvendo plantas, flores e/ou bulbos de espécies ornamentais importados foram realizados e propiciaram informações de inúmeras espécies virais com potencial de causarem perdas de produtividade a grandes culturas, como *Bean yellow mosaic virus* – BYMV (família *Potyviridae*), *Cucumber mosaic virus* – CMV (família *Bromoviridae*) e *Groundnut ringspot virus* – GRSV (família *Bunyaviridae*). Espécies até então não encontradas no país, como o *Alternanthera mosaic virus* – AltMV (*Potexvirus*), foram detectados em material importado. A presença destes patógenos em material importado remete a importância de medidas fitossanitárias e quarentenárias rigorosas para plantas ornamentais (Alexandre et al., 2010).

Novas espécies virais afetando plantas ornamentais também já foram relatadas, como o *Chrysanthemum stem necrosis virus* – CSNV (gênero *Tospovirus* e família *Bunyaviridae*), infectando *Chrysanthemum* sp. (família *Asteraceae*) e espécies de outras famílias botânicas (Bezerra et al. 1999). Em *Dieffenbachia* (família *Araceae*), foram relatados diferentes vírus da família *Potyviridae*, incluindo *Dasheen mosaic virus* – DsMV e outras duas espécies virais não relacionadas a nenhuma espécie descrita até 2003 (Rivas et al., 2003). Uma nova espécie de *Badnavirus*, *Bougainvillea bacilliform virus* (BBV), foi proposta infectando plantas do gênero *Bougainvillea*. Além dessa possível nova espécie, primaveras (*Bougainvillea* sp.) são conhecidas por serem hospedeiras de diferentes espécies virais classificadas nos gêneros *Tospovirus* e *Badnavirus* (Rivas et al., 2005). Em petúnias (*Petunia x hybrida*), Alexandre et al. (2000) caracterizaram um novo tymovírus, *Petunia vein banding virus* – PetVBV.

Vírus em plantas ornamentais podem ou não causar danos significativos às hospedeiras e, portanto, prejuízos aos produtores. Em alguns casos, mesmo quando não há sintomas que comprometam produtividade e/ou aspecto visual para comercialização, a detecção de vírus é de grande importância devido ao potencial reservatório a outras plantas (como grandes culturas). É o caso dos vírus transmitido por *Brevipalpus* (VTB). O complexo conhecido como VTB causa sintomas típicos locais, com anéis cloróticos e pontos verdes em folhas senescentes, independente da espécie. Algumas espécies virais desse complexo são de extrema importância para algumas culturas, como é o caso do *Citrus leprosis virus* (CiLV), *Passion fruit green spot virus* (PFGSV) e *Coffee ringspot virus* (CoRSV). Em um estudo realizado por Freitas-Astua et al. (2010), um grande número de amostras de plantas ornamentais foi analisado visando detectar VTB. Uma ou mais espécies de VTB foram detectadas em 37 espécies de plantas ornamentais de 18 famílias botânicas, incluindo *Araceae*, *Asteraceae*, *Malvaceae*, *Orchidaceae*, *Solanaceae* e outras (Freitas-Astua et al., 2010).

Mesmo com estas pesquisas relatando a presença e infecção por vírus em diferentes plantas ornamentais, estas são, em maioria, baseadas em análises de observação via microscopia

eletrônica e sorologia. Métodos estes que, mesmo convencionais, não apresentam ou direcionam a uma caracterização molecular mais profunda. Quando comparado a grandes culturas, as plantas ornamentais são raramente estudadas e seus vírus pouco conhecidos. Mesmo em grande abundância, os vírus são entidades de difícil estudo e acesso, mediante sua natureza peculiar. Isso, associado a menor importância dada às plantas ornamentais, pode influenciar em potenciais vírus novos ou já conhecidos, porém não sendo detectados e analisados em diferentes espécies de flores e plantas ornamentais.

Relatos de famílias virais que abrigam espécies de grande importância para grandes culturas, hortaliças e fruteiras, como *Potyviridae*, são conhecidas pela ampla gama de hospedeiras e, por isso, podem representar um risco para produtores locais. Outras famílias podem apresentar novas espécies virais em culturas não estudadas, como nas famílias *Luteoviridae* e *Rhabdoviridae*, que ainda apresentam poucos representantes com genoma completamente sequenciado (ICTV, 2016).

3. As famílias virais *Tombusviridae*, *Potyviridae* e *Rhabdoviridae*

3.1. A família *Tombusviridae*

A família *Tombusviridae* atualmente compreende 16 gêneros de vírus. Dentre eles, *Pelarspovirus* e *Umbravirus*, serão apresentados a seguir por corresponderem aos gêneros de duas das prováveis espécies novas virais que serão abordadas neste trabalho. Representantes de *Tombusviridae* possuem RNA de fita simples sentido positivo (ssRNA+) de tamanho variando de 3,7 quilobases (kb) a 4,8 kb. Espécies de *Dianthovirus*, diferem dos demais gêneros por apresentarem genoma bipartido. O critério de distinção entre gêneros na família é complexo, já que alguns representantes possuem características únicas, como a natureza de vírus satélite do gênero *Umbravirus* (Rochon *et al.*, 2009).

De um modo geral, classifica-se gêneros diferentes pela morfologia viral (partículas virais lisas ou granulares), comparação da sequência do genoma (organização, número de segmentos e tamanho do genoma) e características do gene que codifica a polimerase (presença de um códon de terminação ou não)(Rochon *et al.*, 2009). Os diferentes grupos também são distinguidos por comparação de sequências do genoma parcial (polimerase, capsídeo e proteínas de movimento) e completo (Stuart *et al.*, 2004, King *et al.*, 2011).

Pelarspovirus é um gênero recentemente criado e aceito pelo *International Committee on Taxonomy of Viruses* (ICTV), sendo proposto em 2015 (Scheets, Jordan *et al.* 2015). Este é um gênero que abrange espécies com genoma monopartido e apresenta cinco *Open Reading*

Frames (ORFs), incluindo duas ORFs próximas a região 5', duas que codificam proteínas de movimento centrais (MP1 e MP2) e uma que codifica a proteína capsidial (CP) próxima a extremidade 3'. Espécies virais classificadas neste gênero também são distintas ao restante da família por apresentarem a produção de um único RNA subgenômico para expressão dos genes da MP e CP, assim como também na presença de um códon de iniciação (*start codon*) composto apenas de CUG ou GUG (ou seja, total ausência de AUG) nos genes MP1 e CP (Mollov, Lockhart *et al.* 2013). Finalmente, análises filogenéticas já mostravam um grupo distinto, agora separado em um novo gênero – *Pelarspovirus*, entre os antigos membros de *Carmovirus* (agora *Alphacarmovirus*, *Betacarmovirus* e *Gammacarmovirus*). A espécie-tipo de *Pelarspovirus* é *Pelargonium line pattern virus* (PLPV). Os demais integrantes desse grupo são *Elderberry latent virus* (ELV), *Pelargonium chlorotic ring pattern virus* (PCRPV), *Pelargonium ringspot virus* (PeIRSV) e *Rosa rugosa leaf distortion virus* (RrLDV), que foi recentemente relatado em *Rosa rugosa* (Roseira-rugosa) e o genoma completamente sequenciado. Os critérios sugeridos para a demarcação de espécies no gênero *Pelarspovirus* são: identidade menor que 75% de aminoácido entre as polimerases e entre as capas proteicas (Scheets *et al.*, 2015, Mollov *et al.*, 2013, ICTV, 2016).

Vírus classificados no gênero *Umbravirus*, atualmente com sete espécies aceitas como definitivas, apresentam características peculiares quando comparadas às demais espécies da família *Tombusviridae* por serem consideradas como vírus satélite (Talianky & Robinson, 2003). Um vírus de RNA satélite, em sua natureza, não codifica para proteína capsidial necessitando, portanto, de um vírus auxiliar. No caso dos umbravírus, o vírus auxiliar é tipicamente um membro da família *Luteoviridae*. Molecularmente, espécies de *Umbravirus* apresentam um genoma monopartido, linear e de fita positiva. Apresenta quatro ORFs, sendo as ORFs 3 e 4 oriundas de RNAs subgenômicos. A ORF 1 está localizada próxima ao terminal 5' e codifica para uma proteína de 37 kDa. A ORF 2 sobrepõe o final da ORF 1 e codifica uma proteína de 65 kDa, uma polimerase (Rochon *et al.*, 2009). Caso ocorra uma co-infecção com um luteovírus, o genoma do umbravírus pode ser encapsidado com um capsídeo produzido pelo vírus auxiliar. Do contrário, a célula passa a apresentar filamentos de ribonucleoproteínas (Talianky & Robinson, 2003). Para o gênero *Umbravirus*, a demarcação de espécies é realizada com base no círculo de hospedeiros, identidade de nucleotídeos do genoma completo menor que 70% e padrões no dsRNA (quando presentes e não associados ao RNA satélite (Rochon *et al.*, 2009).

A espécie-tipo do gênero *Umbravirus* é *Carrot mottle virus* (CMoV). Este vírus é encontrado em plantas junto com o luteovírus *Carrot red leaf virus* – CRLV (gênero *Luteovirus*

e família *Luteoviridae*). Na hospedeira, cenoura (*Daucus carota*), essa co-infecção está associada aos sintomas de nanismo e mosqueado. Algumas estirpes de CMoV, no entanto, em algumas regiões no mundo, tiveram pouca ou nenhuma relação com os acessos já conhecidos e sequenciados, sugerindo que novas espécies de *Umbravirus*, como o *Carrot mottle mimic virus* (CMoMV), estejam infectando e causando a mesma ou parecida doença (Gibbs *et al.*, 1996).

3.2. A família *Potyviridae*

Atualmente, a segunda maior família de vírus de plantas em termos de número de espécies é a *Potyviridae*, com 195 espécies reconhecidas e distribuídas em oito gêneros a saber: *Brambyvirus*, *Bymovirus*, *Ipomovirus*, *Macluravirus*, *Poacevirus*, *Potyvirus*, *Rymovirus* e *Tritimovirus* (Adams *et al.*, 2009b). Atrás apenas da família *Geminiviridae*, com 362 espécies definitivas (Brown *et al.*, 2009), a família *Potyviridae* consiste, morfológicamente, de partículas virais longas, filamentosas e flexuosas, sendo que o tamanho pode variar entre os diferentes gêneros e espécies. Com exceção das espécies classificadas em *Bymovirus* (genoma bipartido), todas as demais espécies da família apresentam genoma monopartido. Todos os membros de *Potyviridae* apresentam RNA de fita simples senso positivo com a proteína VPg ligada covalentemente ao terminal 5', sendo o terminal 3' poliadenilado. Membros da família apresentam uma ampla gama de hospedeiros, podendo chegar a até 30 famílias botânicas, incluindo dentre suas hospedeiras importantes espécies botânicas de monocotiledôneas e dicotiledôneas. Há, nessa família, vírus que apresentam transmissão por sementes, transmissão mecânica e pelos mais diferentes vetores. Espécies classificadas nos gêneros *Potyvirus*, *Macluravirus* e *Poacevirus* apresentam transmissão por afídeos, enquanto representantes de *Rymovirus* e *Tritimovirus* são transmitidos por ácaros. Espécies de *Ipomovirus* e *Bymovirus* são transmitidos por mosca branca (*Bemisia tabaci*) e fungos, respectivamente. O gênero monotípico, cuja espécie tipo é o *Blackberry virus Y* – BVY não possui vetor conhecido até o momento (Adams *et al.*, 2009b).

Características típicas do grupo incluem a formação de corpos de inclusão citoplasmáticos (e, em alguns casos, nucleares) durante a infecção e a produção de uma única poliproteína, que é seguidamente clivada em uma série de proteínas funcionais (ICTV, 2016, Barnett, 1991, Adams *et al.*, 2005a).

Para a família *Potyviridae*, a demarcação de espécies e gêneros segue critérios descritos por Adams *et al.* (2005). A partir desses critérios, a distinção entre gêneros e espécies é realizado por similaridade genômica, hospedeiro, morfologia dos corpos de inclusão, propriedades antigênicas e diferenças entre os sítios de clivagem na poliproteína.

Especificamente, foi definida uma porcentagem de identidade de nucleotídeo para cada gene, com exceção da P3. O gene que codifica a CP, por exemplo, apresenta, pelo menos, 49% de identidade de nucleotídeo para que os indivíduos pertençam ao mesmo gênero. Em nível de espécie, a taxa de identidade é elevada para 76% (Adams *et al.*, 2005b).

O gênero *Potyvirus*, cuja espécie-tipo é o *Potato virus Y* (PVY), possui 162 espécies, o maior número de espécies descritas até o momento para um gênero dentro de *Potyviridae*. As espécies classificadas neste gênero, apresentam partículas de até 900 nm de comprimento, com a estrutura genômica típica da família, ou seja, com uma única ORF formando uma poliproteína, com cauda Poli (A) e proteína VPg em seu terminal 5' (Adams *et al.*, 2009b). Espécies de *Potyvirus* são conhecidas historicamente por causarem doenças e interferirem drasticamente, reduzindo a produtividade de diferentes culturas, como é o caso da batata (*Solanum tuberosum*) e o PVY. Essa espécie apresenta, também, uma ampla gama de hospedeiras, incluindo espécies fora da família Solanaceae (McDonald & Singh, 1996). Desta forma, a identificação e caracterização de novas espécies do gênero em diferentes plantas, além de contribuir para o desenvolvimento de ferramentas de classificação no gênero, pode significar um avanço na compreensão e biologia de espécies já relatadas e, assim, promover novas formas de controle de doenças mais severas. Para plantas ornamentais, membros do gênero *Potyvirus* já foram identificados em jasmim (*Jasminum* spp.). É o caso de um novo isolado de potyvírus causando anéis cloróticos nas folhas da planta ornamental, denominado de *Jasminum ringspot virus* (JRSV). O vírus foi detectado, identificado e caracterizado molecularmente por meio da técnica de sequenciamento de alto desempenho, ou “*Next Generation Sequencing*” (NGS) aliada a ferramentas de bioinformática (Kaur *et al.*, 2013).

3.3. A família *Rhabdoviridae*

Diferentemente das demais famílias abordadas anteriormente (*Tombusviridae* e *Potyviridae*), a família *Rhabdoviridae* faz parte da ordem Mononegavirales, ou seja, espécies classificadas nos gêneros da família apresentam genoma de RNA senso negativo. Além disso, o grupo é conhecido por ter como membros vírus capazes de infectar invertebrados, vertebrados e plantas, sendo que para as espécies virais que infectam plantas, há uma peculiar morfologia tipo baciliforme típica. Em geral, as espécies apresentam um genoma composto de RNA fita simples negativa (ssRNA-) de tamanho variando de 11 a 15 kb. Podem haver de cinco a seis ORFs e tipicamente os genes são denominados de N, P, M, G e L (codificando o nucleocapsídeo, a fosfoproteína, a proteína de matriz, a glicoproteína e a polimerase, respectivamente) (Dietzgen *et al.*, 2009). Outras ORFs podem ser encontradas, como as ORFs

X e Y da espécie *Potato yellow dwarf virus* (PYDV) e sc4 de *Sonchus yellow net virus* (SYNV) (Jackson *et al.*, 2005). A família *Rhabdoviridae* compreendia, até pouco tempo, apenas seis gêneros, classificados, basicamente, por meio de fatores antigênicos, organização genômica, sítios de replicação e biologia (como gama de hospedeiros e vetores, entre outros). Por meio do desenvolvimento da tecnologia de sequenciamento, novos estudos comparativos e filogenéticos são constantemente realizados e, com isso, hoje, somam-se já 13 gêneros na família, sendo quatro gêneros cujos membros infectam plantas: *Cytorhabdovirus*, *Nucleorhabdovirus*, *Varicosavirus* e o recém aceito *Dichorhavirus* (Dietzgen *et al.*, 2014).

Espécies classificadas em *Cytorhabdovirus* apresentam a característica de se replicarem no citoplasma de células de planta infectadas, geralmente associadas a viroplasmas. Vírions agregam-se ao retículo endoplasmático, acumulando nas vesículas produzidas nessa organela. Além disso, apresentam a morfologia típica baciliforme, podendo chegar a 350 nm de comprimento e genoma composto de um RNA monopartido, senso negativo, como todos os indivíduos da família, e comprimento variando de 12 a 14 kb (Dietzgen *et al.*, 2009). A espécie-tipo é o *Lettuce necrotic yellows cytorhabdovirus* (LNYV), onde o genoma é, à exceção de algumas poucas diferenças, típico para o restante do gênero. Sua sequência apresenta uma ordem de genes da seguinte forma: 3' – N – P – 4b – M – G – L – 5'; respectivamente. Os genes codificam nucleoproteína, fosfoproteína, proteína de movimento, proteína de matriz, glicoproteína e polimerase, respectivamente. Outras características do gênero incluem a presença de uma sequência repetitiva conservada intergênica e a complementariedade entre parte das sequências dos terminais 3' e 5' (ICTV, 2016, Jackson *et al.*, 2005).

Os critérios de demarcação de espécies dentro do gênero *Cytorhabdovirus* seguem a tendência da família: não há um número fixo de limite de identidade genômica, por exemplo. As espécies são, basicamente, diferenciadas pelos hospedeiros e especificidade de vetores. Das dez espécies aceitas como definitivas no gênero, poucas apresentaram seu genoma completamente sequenciado e caracterizado. Isso dificulta uma definição clara na distinção entre espécies a partir de dados de sequências. Esse fato se repete para o restante dos rhabdovírus que infectam plantas (Jackson *et al.*, 2005).

Diferentemente dos *Cytorhabdovirus*, espécies classificadas no gênero *Nucleorhabdovirus* replicam no núcleo das células das plantas hospedeiras. A organela, quando infectada, passa a aumentar consideravelmente em tamanho e apresentar inclusões granulares (viroplasma) (Dietzgen *et al.*, 2009). Tais inclusões são consideradas os sítios de replicação do vírus. Sua morfologia é similar aos *Cytorhabdovirus*, com partículas baciliformes podendo chegar até 300 nm de comprimento. Os genomas, da mesma forma, também seguem o padrão citado para a

família, codificam de seis a sete ORFs, basicamente, na seguinte ordem: 3' – N – X – P – Y – M – G – L – 5'. Ainda não se conhece a função do gene X (Jackson, Dietzgen *et al.* 2005). Em alguns casos, como para *Sonchus yellow net nucleorhabdovirus* (SYNV), há o gene sc4, que possui a função de movimento (além de possíveis outras ainda não bem determinadas) (Wang, Ma *et al.* 2015). Espécies de *Nucleorhabdovirus* também apresentam sequências padrões/repetitivas intergênicas (Jackson *et al.*, 2005, Wang *et al.*, 2015).

A espécie-tipo do gênero *Nucleorhabdovirus* é *Potato yellow dwarf nucleorhabdovirus* (PYDV) que foi primeiramente estudada em associação a uma doença grave em batata. Esse estudo envolvendo uma espécie de rhabdovírus e a hospedeira batata, realizado a partir da década de 1920, contribuiu para o entendimento das relações e interações vírus-inseto vetor, além de definir a estrutura e morfologia de rhabdovírus de planta em geral (Reeder *et al.*, 1972). Apesar dessa importância, o genoma de PYDV só foi totalmente sequenciado há pouco tempo, em 2008 (Jackson, Dietzgen *et al.* 2005). Estudos comparativos avaliaram as diferenças observadas em células de plantas infectadas entre o PYDV e outro *Nucleorhabdovirus* amplamente estudado, o SYNV. Além das diferenças no comportamento de algumas hospedeiras infectadas com os dois vírus separadamente, viu-se que o PYDV altera o envelope nuclear da célula do hospedeiro a partir da periferia do núcleo. Comparativamente, o SYNV induz acumulação de membranas nucleares no interior do próprio núcleo. Essa avaliação citopatológica contribui para a diferenciação destes dois nucleorhabdovírus próximos geneticamente (Jackson *et al.*, 2005).

Sonchus yellow net nucleorhabdovirus (SYNV) é a espécie mais estudada entre os rhabdovírus que infectam plantas. É um vírus transmitido por afídeos e infecta plantas daninhas das espécies *Sonchus oleraceus* e *Bidens pilosa*, além de, esporadicamente, alface (*Lactuca sativa*) (Scholthof *et al.*, 1994). Suas partículas seguem o padrão da família e gênero, porém o seu genoma, além dos típicos N, M – este dividido em M1 e M2, G e L e apresenta um gene característico, o sc4 (Goodin *et al.*, 2001). O gene sc4 apresenta 1196 nucleotídeos, com uma ORF de 972 nucleotídeos capaz de codificar uma proteína de 324 aminoácidos. Assim como os demais genes, o sc4 também é circundado de sequências repetitivas conservadas na região intergênica. Acredita-se que a função do sc4 seja de proteína de movimento, além da associação com a glicoproteína (G) na membrana fosfolipídica (Wang *et al.*, 2015, Goodin *et al.*, 2001, Scholthof *et al.*, 1994).

Lettuce big-vein varicosavirus (LBVV) é a espécie-tipo (e, por enquanto, única aceita) do gênero *Varicosavirus* e apresenta genoma bipartido, uma exceção aos demais rhabdovírus juntamente com os vírus classificados no gênero *Dichoravirus* (Dietzgen *et al.*, 2014). A

espécie LBVV apresenta, além do genoma de natureza bipartida, outras características peculiares. A partícula do LBVV não é envelopada e a transmissão no solo é feita por zoósporos de fungos vetores, além de as partículas virais permanecerem de maneira estável em esporos destes mesmos fungos. Essa relação patógeno-vetor dificulta o controle da doença causada pelo vírus (Sasaya *et al.*, 2002).

Orquídeas com sintomas de clorose e manchas necróticas foram avaliadas no Japão em 1977. Logo se caracterizou o agente etiológico: um vírus, denominado *Orchid fleck virus* (OFV). Posteriormente, foi relatada a presença de OFV em diferentes regiões ao redor do globo em diferentes espécies de orquídeas e todas apresentando sintomas similares (Dietzgen *et al.*, 2014). Apesar de ser experimentalmente transmitido via transmissão mecânica para plantas da família Orchidaceae e outras famílias, sabe-se que o OFV tem a sua maior disseminação ao ácaro vetor *Brevipalpus* spp. (Freitas-Astua *et al.*, 2010). Uma das características desse rhabdovírus é o seu genoma bipartido, consistindo de duas moléculas, RNA1 e RN2. Vírus também transmitidos pelo mesmo vetor, como *Coffee ringspot virus* (CoRSV) e *Citrus leprosis virus* (CiLV), também apresentam o genoma composto de RNA1 e RNA2. Ainda, as três espécies virais também não apresentavam membrana lipídica. Assim, mesmo em contraposição à estrutura da ordem Mononegavirales, que estipula apenas vírus monopartidos, e da família, que até então apresentava todos os seus representantes como envelopados, criou-se um novo gênero dentro de *Rhabdoviridae*: *Dichorhavirus*, com o *Orchid fleck virus* (OFV) sendo a espécie-tipo. Os dichorhavirus, mesmo tão únicos perante o restante da família e até mesmo da ordem, apresentam sequências com alta identidade com o gênero *Nucleorhabdovirus* nos genes G e L. Mesmo assim, considera-se possível que as espécies de *Dichorhavirus* possam, em algum momento, serem classificadas no gênero, porém sem família ou ordem definidas (Dietzgen *et al.*, 2014).

Por apresentarem uma morfologia característica, há uma relativa facilidade na distinção entre células infectadas por vírus da família *Rhabdoviridae* e saudáveis, quando visualizados em microscopia eletrônica de transmissão. Desta forma, muitos membros novos foram descritos infectando diferentes hospedeiras, mesmo sem confirmação por meio de ensaios moleculares (Freitas-Astua *et al.*, 2010). Há, portanto, a necessidade de se analisar molecularmente e filogeneticamente tais espécies (Jackson *et al.*, 2005). Nesse sentido, análises moleculares e filogenéticas estão sendo empregadas em grupos já estabelecidos, resultando nestes sendo reordenados ou refeitos. É o caso do novo gênero *Dichorhavirus* e de alguns de suas espécies, como o CiLV. Este, por exemplo, passou a ser estirpe da espécie tipo do gênero, o OFV, por apresentarem alta identidade em suas ORFs (Dietzgen *et al.*, 2014). Atualmente, essa

problemática tem sido rebatida por meio das novas técnicas de sequenciamento, que facilitam a análise e caracterização de novos vírus e vírus já caracterizados (Dietzgen *et al.*, 2014).

4. Sequenciamento de alto desempenho (*Next Generation Sequencing* – NGS)

A genômica, até pouco tempo, era dependente de alguns métodos de sequenciamento, entre eles a técnica enzimática de Sanger, descrita em 1977. O método de Sanger foi usado no primeiro projeto de fluorescência automatizada para sequenciar uma região de um genoma. Nesse projeto, houve a introdução, também, do conceito de sequenciamento em “*paired-end*”. Com o sucesso do projeto e a região do genoma sequenciada com êxito e sem grandes ambiguidades, surgiu a noção da capacidade de grandes sequenciamentos utilizando técnicas com base em fluorescência automatizada (Ansorge, 2009). Com o movimento da comunidade internacional acerca do projeto Genoma Humano, houve uma corrida para o desenvolvimento de novas técnicas que pudessem aprimorar o desempenho de sequenciamento. Observou-se que um dos grandes limites para a plataforma Sanger era a necessidade de gel ou polímeros para a separação física dos fragmentos de DNA, além da quantidade baixa de amostras que são analisadas em paralelo. Assim, focou-se o estudo para o desenvolvimento de novas técnicas e plataformas que não necessitassem de gel, aumentando o número de sequências em milhões. Neste contexto e com o auxílio de novas técnicas dispostas à época, surgiram as novas plataformas que, hoje, entendemos como de “alto desempenho”, “alta eficiência” ou, simplesmente, “nova geração” - NGS (Ansorge, 2009).

O conceito de NGS é relativamente similar ao de sequenciamento via Sanger. Bases de um fragmento de sequência são identificados em sequência e, em seguida, resintetizados de um DNA molde. A tecnologia do NGS permite multiplicar este processo milhares de vezes, em um sistema onde incontáveis reações ocorrem paralelamente. Equipamentos e ferramentas modernas permitem, hoje, produzir centenas de gigabases em dados em uma única corrida, tornando possível sequenciar genomas inteiros em poucos passos e dias (von Bubnoff, 2008). O método envolve a formação e o sequenciamento de fragmentos de sequências (*reads*) a partir de uma amostra, uma extração total de DNA ou RNA, por exemplo.

De acordo com Adams *et al.* (2009), NGS é uma potencial ferramenta de diagnose em doenças de plantas. Neste trabalho, buscou-se identificar um vírus que causava mosaico em *Liatris spicata*. Não havendo qualquer informação quanto ao agente, a amostra foi aplicada (cDNA a partir de extração de RNA de uma planta indicadora, *Gomphrena globosa*) em um sequenciador de “nova geração”, gerando centenas de milhares de dados. No total, foram 71.146

reads formados, com uma média de 227 pb para cada, resultando em um total de 16.149.138 pb sequenciados. Examinando tais informações via BLAST, os autores verificaram que a maioria (de origem provável viral) era semelhante, mas não igual, a três cucumovirus: *Cucumber mosaic virus* (CMV), *Peanut stunt virus* (PSV) e *Tomato aspermy virus* (TAV). Por interpretação e aplicação bioinformática, produziu-se *contigs* a partir dos inúmeros *reads*, resultando em três principais segmentos de genoma: seria, portanto, um cucumovírus novo tripartido, o *Gayfeather mild mottle virus* (Adams *et al.*, 2009a).

Pesquisas envolvendo detecção e identificação de patógenos em plantas utilizam diversos métodos. A maioria, porém, envolve uma predição de qual agente estaria envolvido no processo, aplicando métodos específicos ao patógeno-candidato. Isso pode limitar a descoberta de novos patógenos e, se aquele candidato não estiver presente, comprometer os resultados. Este tipo de metodologia fica refém de dados publicados na literatura, o que pode prejudicar especialmente alguns tipos de hospedeiras, como as ornamentais, por exemplo. Métodos que envolvem a montagem *de novo* de sequências são uma forma de superar este empecilho: sequenciamento via NGS é um exemplo de metodologia moderna que é capaz de gerar tais dados, cabendo a análise a ser realizada por ferramentas de bioinformática analisá-los (Adams *et al.*, 2009a).

O NGS, pela sua complexidade e eficiência na geração de dados, configura-se como um versátil método para diversos fins: diagnose, metagenômica, estudo populacional, entre outros. A identificação de novos vírus era, até pouco tempo atrás, motivo de estudos envolvendo extensas equipes, trabalho oneroso e longos períodos de pesquisa. Com o advento da tecnologia de sequenciamento, a montagem de genomas via *de novo* compreende um passo importante para a fitopatologia e microbiologia em geral.

As tecnologias do NGS se baseiam em diferentes plataformas de sequenciamento, de diferentes estratégias e empresas. As mais populares são a *Roche Genome Sequencer FLX* (FLX), *Applied Biosystems SOLiD* (SOLiD) e *Illumina Genome Analyzer* (GA). Essas tecnologias se mostraram superiores ao sequenciamento clássico Sanger por gerarem um número maior de dados em um custo razoável para a proporção. Além disso, são capazes de atuar em diferentes áreas, como transcriptoma, análise de DNA metilado e, naturalmente, metagenômica. Como a característica do NGS é gerar milhões de dados, primariamente *reads*, deve-se imaginar que há diferenças no número de sequências e de qualidade nestas entre uma tecnologia e outra. Isso é importante, pois remete diretamente aos posteriores procedimentos de bioinformática. Em relação ao volume bruto de dados gerados, a plataforma SOLiD conseguiu alcançar o patamar de 226 milhões de *reads* em uma reação de sequenciamento NGS para a

bactéria *Escherichia coli*. Em seguida, destaca-se a GA, com 9 milhões, e a FLX, com 475 mil. A diferença abrupta entre a capacidade de FLX e demais em gerar dados ocorre, principalmente, pelo tamanho das sequências que gera – algo em torno de 260 pb. Na GA, a sequência atinge 36 pb, enquanto que na SOLiD, pode chegar a 50. A redundância de informação na SOLiD é alta, resultando em baixo valor de qualidade. O grau de fidelidade das sequências, por sua vez, é próxima entre as três plataformas, indicando que, independente de uma série de fatores, as tecnologias de NGS prevaleceram como altamente confiáveis (Suzuki *et al.*, 2011). A escolha da plataforma leva em conta os custos e recursos disponíveis. Tipicamente, compara-se o investimento em tecnologias NGS com o sequenciamento Sanger, pela alta diferença entre os métodos. Enquanto que a FLX apresenta um gasto 10 vezes menor que o Sanger para sequenciar um genoma, a GA, da *Illumina*, apresenta um valor de 100 vezes menor (von Bubnoff, 2008).

Novas tecnologias podem revolucionar o método científico em qualquer área. Assim foi na virologia com a microscopia eletrônica de transmissão, a *Polimerase Chain Reaction* (PCR) e, atualmente, com o NGS (Mokili *et al.*, 2012). Estas metodologias inovadoras mudam a forma de como os estudos relacionados a vírus são feitos, particularmente em áreas como sequenciamento, estudo da evolução, ecologia, transcriptômica e metagenômica. Nessa última, o NGS tem contribuído para a aceleração da descoberta de novos vírus (Mokili *et al.*, 2012, Radford *et al.*, 2012).

Com o advento da tecnologia, obter e gerar centenas de milhares de dados é, hoje, considerado uma tarefa relativamente fácil. Os ensaios posteriores, notadamente os de bioinformática, são criticamente importantes no processo de análise de dados e geração de resultados concisos. Obter e analisar os dados tornou-se automatizado, o que pode comprometer a qualidade na interpretação destes. Estudar, organizar e interpretar tamanho número de dados gerados carece de um estudo prévio aprofundado e uma noção em técnicas e métodos em bioinformática (Radford *et al.*, 2012).

A tecnologia do sequenciamento NGS permite a detecção, identificação e descoberta de novos vírus em plantas de uma maneira direta e sem conhecimento prévio do agente ou da doença. Assim, torna-se possível ter noções de populações virais em um determinado ecossistema ou grupo de plantas selecionadas. A tecnologia tem a vantagem de atuar em casos onde não há sequências de referência, porém a falta de uma cobertura significativa prévia pode acarretar em possíveis erros ao longo da interpretação de dados (Prabha *et al.*, 2013)

5. Bioinformática aplicada a virologia vegetal

Com o advento da tecnologia do NGS, é possível sequenciar milhões de pares de base em questão de horas. Desta forma, torna-se possível e acessível o sequenciamento em larga escala de ácidos nucleicos provenientes de diferentes ambientes ou organismos, método denominado metagenômica. As tecnologias atuais, no entanto, geram sequências pequenas (*reads*), de 35 a 100 pares de base, dependendo da metodologia. Esse tipo de informação gerada causa um transtorno no momento de análise, anotação e construção de genomas. A dificuldade imposta pelo número de *reads* não só remete às questões técnicas de bioinformática, mas na quantidade e tamanho dos arquivos de informação, necessitando uma atenção a mais ao armazenamento. Há, portanto, esses distintos pontos que, ainda, são obstáculos para os trabalhos envolvendo NGS e Metagenômica (Pop & Salzberg, 2008).

Metagenômica é um estudo que envolve análises de microrganismos em diferentes ambientes por meio de sequenciamento. Com o advento do NGS e, conseqüentemente, do sequenciamento de alto desempenho por um preço mais acessível, a metagenômica atingiu seu ápice em estudos em plantas (Stobbe & Roossinck, 2014). Com as plantas como um ambiente a ser analisado, é possível obter dados referentes a fungos, bactérias e vírus. Essa estratégia é especialmente interessante para o diagnóstico de fitopatógenos, auxiliando o entendimento da comunidade microbiana como um todo (Roossinck *et al.*, 2015). Por meio desse uso das novas tecnologias, foi possível questionar se vírus são mesmo intrínsecos a patologias em plantas. Com os estudos de metagenômica surgindo, novos dados referentes a distintos táxons de vírus foram aparecendo e, em alguns casos, estes microrganismos não estavam associados a nenhuma doença. Com os mesmos estudos, no entanto, viu-se os resultados da globalização: a circulação de diferentes espécies virais entre países, o que pode gerar potenciais riscos para novas interações vírus-hospedeiro, tanto benéficas como maléficas (Stobbe & Roossinck, 2014).

A metagenômica consiste em um estudo da diversidade ecológica de organismos em determinado meio utilizando a biologia molecular. Muitos microrganismos não são passíveis de serem cultivados em meio de cultura; na verdade, presume-se que apenas 0,1 a 1% do total da microbiota tenha sido isolada em meio até hoje. Neste grupo estão os vírus de plantas, parasitas intracelulares obrigatórios por excelência e, portanto, impossíveis de serem isolados em cultura (Hull, 2013). A obtenção de informações genéticas, biológicas e populacionais destes organismos, sejam vírus, bactérias, archaea, fitoplasmas, protozoários, entre outros, consiste em um fator primordial para a definição de novos métodos de controle para doenças em culturas de importância econômica. Um estudo mais abrangente envolvendo este ramo da

biologia, englobando toda a ecologia de determinado ambiente, é o alicerce da metagenômica. Este processo envolve a confecção de bibliotecas genômicas de uma gama de microrganismos de um meio, sendo os resultados obtidos com a metagenômica um reflexo da complexidade que é o meio e a biota que lhe é compreendida (Lara-Victoriano *et al.*, 2011).

Para vírus de plantas, a metagenômica vem sendo realizada utilizando inicialmente técnicas que visam enriquecimento viral previamente ao sequenciamento, por meio de: semi-purificação de partículas virais seguida de extração de ácido nucleico, DNA ou RNA, extração de dsRNA (double strand RNA – RNA de fita dupla) e sRNA (small RNA – pequenos RNAs produzidos em decorrência de mecanismos de defesa de plantas a vírus, como silenciamento gênico). Para vírus de DNA, utiliza-se, ainda, RCA (*Rolling Circle Amplification*). Posteriormente, o material é enviado para sequenciamento. Após esta etapa, com as regiões sequenciadas, genes podem ser avaliados e caracterizados (Roossinck *et al.*, 2015). As aplicações da metagenômica variam, mas seu uso como fonte de diagnose na fitopatologia tem ganhado força com o advento de novas tecnologias, como a do NGS (Lara-Victoriano *et al.*, 2011).

A metagenômica para vírus de plantas significou um grande avanço para a descoberta de novos vírus, além de um melhor entendimento em relação aos genomas em geral. Observou-se com estas novas sequências virais surgindo a todo momento a dimensão da diversidade viral presente em plantas. Entre 60 a 95% destas sequências não apresentam identidade com sequências de bancos de dados virais, o que pode representar, ou não, novos grupos de vírus. Enquanto a metodologia típica de metagenômica utiliza amostras ambientais, como água de esgoto, fezes, solo, entre outros, muitos estudos agora voltam para amostras individuais de plantas. Isso faz com que seja possível analisar e identificar diferentes estruturas ecológicas virais dentro do hospedeiro, já que os resultados gerados pela metagenômica estariam especificados a um único ambiente, a planta hospedeira. Apesar de em muitos casos os vírus identificados não apresentarem qualquer tipo de associação a doença com o hospedeiro trabalhado, isso não impede que o mesmo venha a infectar diferentes plantas/culturas no momento ou em um futuro próximo (Roossinck *et al.*, 2015).

O advento simultâneo das novas tecnologias de sequenciamento e das técnicas de metagenômica resultaram em um aumento massivo na quantidade de informações e de sequenciamento. Esse tipo de resultado aproxima os estudos do objetivo principal da metagenômica: a identificação, análise e sequenciamento de todos os genomas dentro de uma comunidade/ambiente/amostra. Ao mesmo tempo, no entanto, gera também grandes

dificuldades e desafios que devem ser tomados em conta para que os resultados significativos acompanhem o fluxo de novas informações (Scholz *et al.*, 2012).

A quantidade de informação gerada a cada sequenciamento é extremamente grande quando comparada às tecnologias pré-NGS. Gerir esse tipo de dado está cada vez mais complexo, visto o aumento no número de sequenciamentos e trabalhos envolvendo metagenômica. Com a complexidade, perde-se boa parte das informações. Este “gargalo” é, atualmente, um dos principais temas da comunidade envolvida nesse tipo de sequenciamento. Esse gargalo é, basicamente, a incapacidade de se lidar com o número de dados disponíveis para cada amostra. Em muitos casos, ainda, vê-se o desperdício de dados nos estudos. Enquanto o sequenciamento gera centenas de milhões de *reads*, poucos (centenas) são usados para o posterior trabalho de análise e construção de genoma. Isso pode resultar em falsos-negativos e resultados superficiais (Scholz *et al.*, 2012).

Apesar de novas ferramentas terem sido desenvolvidas especificamente para evitar o gargalo, a contínua produção de novas sequências e dados demanda novos algoritmos periodicamente. Para evitar que o volume de dados não sobrecarregue os recursos dispostos atualmente, cabe, portanto, aos bioinformatas em tentar reverter a situação, com o desenvolvimento de novos softwares, ferramentas, algoritmos e estratégias. Esse ramo, hoje, dispõe e tornou possível as principais metodologias envolvendo a análise de dados gerada por NGS (Scholz *et al.*, 2012).

A bioinformática é uma área que participa em diversos ramos da Biologia. Para a genômica e o NGS, a bioinformática passou a estudar questões como interpretação e armazenamento de dados, controle de qualidade e outros. Com a nova demanda, também surgiram novos pesquisadores capacitados para a análise computacional exigida pelas novas tecnologias de sequenciamento (Weier, 2013). Dentro da área da genômica, os vírus foram os primeiros sistemas biológicos completamente sequenciados, como o bacteriófago MS2, em 1976. Na década de 1990, alguns vírus patogênicos ao homem foram completamente sequenciados da mesma forma, em especial o *Human immunodeficiency virus* – HIV (*Retroviridae*). Após isso, foi obtido o êxito no sequenciamento do genoma humano. A partir desse momento, no entanto, as atenções se voltaram aos organismos mais complexos, quando genomas virais passaram a ser pouco trabalhados e técnicas computacionais deixaram de ser desenvolvidas exclusivamente para tais finalidades. Esse aspecto mudou significativamente apenas com o NGS, em que a facilidade e o preço acessível fizeram com que se passasse a descobrir novos genomas periodicamente (Marz *et al.*, 2014).

O fluxo de trabalho típico para o bioinformata na área de virologia vegetal envolve análise de qualidade dos sequenciamentos, organização, montagem dos genomas e anotação dos mesmos. Cada etapa requer diferentes metodologias e ferramentas próprias, além de recursos computacionais consideráveis. Além disso, uma etapa é dependente da outra. Uma vez os *reads* não sendo analisados qualitativamente e as regiões de baixa qualidade não cortadas (*trimmadas*), o restante do processo de montagem de uma sequência de consenso poderá resultar em análises pouco satisfatórias. Desta forma, *softwares* para estes fins específicos são primordiais para pesquisas envolvendo NGS. Nesse sentido, programas como o *Trimmomatic* apresentam a capacidade de lidar e excluir regiões de *reads* onde a qualidade (a ser determinada pelo operante) está abaixo do desejado (Bolger *et al.*, 2014).

Seguindo o fluxo de trabalho, os *reads* já analisados e trabalhados, ou seja, sem as regiões de baixa qualidade de sequência, podem ser agrupados em montadores (em inglês “*assemblers*”), formando os *contigs*. A forma como os *contigs* são montados varia imensamente dependendo do tipo de montador utilizado. Atualmente, diversos são os softwares disponíveis, porém a maioria é destinada a algum organismo ou tipo de ambiente. Uma grande parte dos programas de montadores é destinada a sequências de bactérias, ou seja, seus parâmetros são pré-configurados para sequências consideravelmente maiores que as virais. Há, no entanto, formas de lidar com esse obstáculo, alterando os parâmetros e variando os montadores até se obter os resultados considerados ótimos. Alguns montadores são continuamente atualizados e, hoje, podem até lidar com sequências obtidas via metagenômica, como é o caso do *software SPAdes* (Bankevich *et al.*, 2012). Outros são específicos para dados de sequenciamento com *reads* pequenos, como os programas *Velvet* (Zerbino & Birney, 2008) e *ABYSS* (“*Assembly By Short Sequences*”) (Simpson *et al.*, 2009).

Para a virologia vegetal, assim como para outras áreas, não há nenhum protocolo de bioinformática a ser seguido para um trabalho envolvendo NGS. Diferentes montadores e demais *softwares* podem ser usados de forma conjunta, de maneira a se obter resultados próximos e condizentes um com o outro. Uma vez com estes, cabe ao pesquisador optar por um ótimo e utilizar nos trabalhos. Uma vez com os *contigs*, o pesquisador poderá montar seu genoma até obtê-lo de maneira completa ou parcial, a depender das condições de amostragem e sequenciamento. Sabe-se, no entanto, que a montagem e posterior análise dos genomas montados requer uma visualização que transcende a janela de comando do computador. A noção visual e espacial permite que o usuário possa interpretar os dados de forma mais consolidada. Desta forma, *softwares* tipo “plataforma” garantem este apelo visual para as análises de bioinformática, dando maior segurança até para usuários com pouca aptidão

computacional. O *software Geneious* (Biomatters Limited) é um exemplo, viabilizando a apresentação das sequências de forma visual que permite analisar, montar e fazer anotações de genomas, sequências de proteínas e outros (Kearse *et al.*, 2012). Com esses exemplos, aponta-se que o principal objetivo da bioinformática nesse momento de novos genomas surgindo periodicamente é a facilitação do uso dos dados obtidos do sequenciamento para geração de sequências confiáveis e mais completa possível (Kearse *et al.*, 2012, Pop & Salzberg, 2008).

6. Referências Bibliográficas

Abafep, 2014. Associação Brasileira do Agronegócio de Flores e Plantas. In. (2015.)

Adams IP, Glover RH, Monger WA, *et al.*, 2009a. *Next-Generation Sequencing* and metagenomic analysis: a universal diagnostic tool in plant virology. *Molecular Plant Pathology* **10**, 537-45.

Adams MJ, Antoniw JF, Beaudoin F, 2005a. Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. *Molecular Plant Pathology* **10**, 537-45.
6, 471-87.

Adams MJ, Antoniw JF, Fauquet CM, 2005b. Molecular criteria for genus and species discrimination within the family Potyviridae. *Archives of Virology* **150**, 459-79.

Adams MJ, Zerbini FM, French R, Rabenstein F, Stenger DC, Valkonen JPT, 2009b. *Virus Taxonomy: Ninth Report - Potyviridae family*. ICTV.

Alexandre MaV, Duarte LML, Rivas EB, Galleti SR, 2010. Vírus detectados em plantas ornamentais no período 2004 a 2008. *Revista Brasileira de Horticultura Ornamental* **16**, 95-100.

Alonso AM, Silva JCS, 2009. Floricultura no Distrito Federal: perspectivas para o fortalecimento da cadeia produtiva de flores e plantas ornamentais. In. Planaltina, DF: Embrapa Cerrados. (2015.)

Ansorge WJ, 2009. Next-generation DNA sequencing techniques. *New Biotechnology* **25**, 195-203.

Baker KF, Linderman RG, 1979. Unique features of the pathology of ornamental plants. *Annual Review Phytopathology* **17**, 25.

Bankevich A, Nurk S, Antipov D, *et al.*, 2012. *SPAdes*: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**, 455-77.

Barnett OW, 1991. Potyviridae, a proposed family of plant viruses. *Archives of Virology* **118**, 139-41.

- Bolger AM, Lohse M, Usadel B, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20.
- Brown JK, Fauquet CM, Briddon RW, Zerbini M, Moriones E, Navas-Castillo J, 2009. *Virus Taxonomy: Ninth Report - Geminiviridae family*. ICTV.
- Dietzgen RG, Calisher CH, Kurath G, *et al.*, 2009. *Virus Taxonomy: Ninth Report - Rhabdoviridae family*. ICTV.
- Dietzgen RG, Kuhn JH, Clawson AN, *et al.*, 2014. Dichorhavirus: a proposed new genus for Brevipalpus mite-transmitted, nuclear, bacilliform, bipartite, negative-strand RNA plant viruses. *Archives of Virology* **159**, 607-19.
- Freitas-Astua J, Rodrigues JC, Kitajima EW, 2010. An annotated list of ornamentals naturally found infected by Brevipalpus mite-transmitted viruses. *Scientia Agricola* **67**, 348-71.
- Gibbs MJ, Ziegler A, Robinson DJ, Waterhouse PM, Cooper A, 1996. Carrot mottle mimic virus (CMoMV): A second umbravirus associated with carrot motley dwarf disease recognised by nucleic acid hybridisation. In. *Molecular Plant Pathology On-Line*. (2017.)
- Goodin MM, Austin J, Tobias R, Fujita M, Morales C, Jackson AO, 2001. Interactions and nuclear import of the N and P proteins of sonchus yellow net virus, a plant *Nucleorhabdovirus*. *Journal of Virology* **75**, 9393-406.
- Hull R, 2013. *Plant Virology*. Elsevier.
- Ibraflor, 2014. Instituto Brasileiro de Floricultura. In. (2015.)
- Jackson AO, Dietzgen RG, Goodin MM, Bragg JN, Deng M, 2005. Biology of plant rhabdoviruses. *Annual Review Phytopathology* **43**, 623-60.
- Junqueira AH, Peetz MS, 2008. Mercado interno para os produtos da floricultura brasileira: características, tendências e importância socioeconômica recente. *Revista Brasileira de Horticultura Ornamental* **14**, 37-52.
- Junqueira AH, Peetz MS, 2015. Flores e Plantas ornamentais do Brasil - Série estudos mercadológicos. In: Sebrae SBDa-ÀMEPE-, ed. Brasília - DF: SEBRAE, 1-44. (1.)
- Kaur C, Kumar S, Snehi SK, Raj SK, 2013. Molecular detection of Jasmine potyvirus associated with yellow mosaic symptoms on Jasminum sambac L. in India. *Plant Molecular Virology* **46**, 8.
- Kearse M, Moir R, Wilson A, *et al.*, 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-9.
- King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, 2011. *Ninth Report of the International Committee on Taxonomy of Viruses: Classification and Nomenclature of Viruses*. Elsevier.

- Lara-Victoriano F, Castillo-Reyes F, Flores-Gallegos C, Aguilar CN, Herrera-Rodríguez R, 2011. Metagenomics in plant pathology. *Phytopathology in the Omics Era* **2**, 9.
- Marz M, Beerenwinkel N, Drosten C, *et al.*, 2014. Challenges in RNA virus bioinformatics. *Bioinformatics* **30**, 1793-9.
- Mcdonald JG, Singh RP, 1996. Host Range, Symptomology, and Serology of Isolates of Potato virus Y (PVY) That Share Properties with Both the PVYn and PVYo Strain Groups. *American Potato Journal* **73**, 7.
- Mokili JL, Rohwer F, Dutilh BE, 2012. Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology* **2**, 63-77.
- Molloy D, Lockhart B, Zlesak DC, 2013. Complete nucleotide sequence of Rosa rugosa leaf distortion virus, a new member of the family Tombusviridae. *Archives of Virology* **158**, 2617-20.
- Paiva PD, 2008. *Paisagismo: Conceitos e Aplicações*. Lavras: UFLA.
- Pop M, Salzberg SL, 2008. Bioinformatics challenges of new sequencing technology. *Trends Genetics* **24**, 142-9.
- Prabha K, Baranwal VK, Jain RK, 2013. Applications of next generation high throughput sequencing technologies in characterization, discovery and molecular interaction of plant viruses. *Indian Journal of Virology* **24**, 157-65.
- Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N, 2012. Application of *Next-Generation Sequencing* technologies in virology. *Journal of General Virology* **93**, 1853-68.
- Reeder GS, Knudson DL, Macleod R, 1972. The ribonucleic acid of potato yellow dwarf virus. *Virology* **50**, 301-4.
- Rivas EB, Duarte LML, Alexandre MaV, Fernandes FMC, Harakava R, Chagas CM, 2005. A new Badnavirus species detected in Bougainvillea in Brazil. *Journal of General Plant Pathology* **71**, 3.
- Rivas EB, Galletti SR, Duarte LML, Alexandre MaV, Cilli A, Estelita MEM, 2003. Detecção de Potyviridae em espécies de Dieffenbachia. *Arquivos do Instituto Biológico* **70**, 3.
- Rochon D, Lommel S, Martelli GP, Rubino L, Russo M, 2009. *Virus Taxonomy: Ninth Report - Tombusviridae family*. ICTV.
- Roossinck MJ, Martin DP, Roumagnac P, 2015. Plant Virus Metagenomics: Advances in Virus Discovery. *Phytopathology* **105**, 716-27.
- Ruud LM, Nico SP, 2005. Floriculture World Wide; production, trade and consumption patterns show market opportunities and challenges. In. Wageningen University and Research centre. The Netherlands.: Wageningen University and Research centre. (2015.)

- Sasaya T, Ishikawa K, Koganezawa H, 2002. The nucleotide sequence of RNA1 of Lettuce big-vein virus, genus Varicosavirus, reveals its relation to nonsegmented negative-strand RNA viruses. *Virology* **297**, 289-97.
- Scheets K, Jordan R, White KA, Hernandez C, 2015. Pelarspovirus, a proposed new genus in the family Tombusviridae. *Archives of Virology* **160**, 2385-93.
- Scholthof KB, Hillman BI, Modrell B, Heaton LA, Jackson AO, 1994. Characterization and detection of sc4: a sixth gene encoded by sonchus yellow net virus. *Virology* **204**, 279-88.
- Scholz MB, Lo CC, Chain PS, 2012. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Current Opinion in Biotechnology* **23**, 9-15.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I, 2009. ABySS: a parallel assembler for short read sequence data. *Genome Research* **19**, 1117-23.
- Stobbe AH, Roossinck MJ, 2014. Plant virus metagenomics: what we know and why we need to know more. *Frontiers in Plant Science* **5**, 150.
- Stuart G, Moffett K, Bozarth RF, 2004. A whole genome perspective on the phylogeny of the plant virus family Tombusviridae. *Archives of Virology* **149**, 1595-610.
- Suzuki S, Ono N, Furusawa C, Ying BW, Yomo T, 2011. Comparison of sequence reads obtained from three Next-Generation Sequencing platforms. *PLoS One* **6**, e19534.
- Taliansky ME, Robinson DJ, 2003. Molecular biology of umbraviruses: phantom warriors. *Journal of General Virology* **84**, 1951-60.
- Uncomtrade, 2014. UN Comtrade Database. In. (2015.)
- Von Bubnoff A, 2008. Next-Generation Sequencing: the race is on. *Cell* **132**, 721-3.
- Wang Q, Ma X, Qian S, *et al.*, 2015. Rescue of a Plant Negative-Strand RNA Virus from Cloned cDNA: Insights into Enveloped Plant Virus Movement and Morphogenesis. *PLoS Pathogens* **11**, e1005223.
- Weier HG, 2013. Bioinformatics for High Throughput Sequencing. *Data Mining in Genomics & Proteomics* **4**, 3.
- Zerbino DR, Birney E, 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**, 821-9.

CAPÍTULO 2

VIROMA EM ESPÉCIES ORNAMENTAIS PRODUZIDAS NO DISTRITO FEDERAL

RESUMO

Amostras de dez plantas ornamentais produzidas no Viveiro I da Companhia Urbanizadora da Nova Capital (NOVACAP) em Brasília, Distrito Federal, foram coletadas apresentando sintomas típicos de infecção por vírus, como mosaico, mosqueado, amarelecimento, encarquilhamento, anéis cloróticos e ápice encarquilhado. Para a detecção de vírus nas dez amostras (*Pinanga kuhlii*, *Anthurium lindmanianum*, *Impatiens hawkeri*, *Pelargonium* sp., *Epipremnum aureum*, *Jasminum nitidum*, *Coreopsis lanceolata*, *Pachystachys lutea*, *Neomarica candida* e *Streptosolen jasmonii*) coletadas, um *pool* composto foi formado. Após enriquecimento viral mediante semi-purificação de partículas virais e extração de RNA, uma única amostra composta foi submetida ao *Next Generation Sequencing* (NGS), que resultou em 54.209.722 *reads*. Estes *reads* foram analisados por ferramentas de bioinformática até se obter o máximo de 2.882.857 *contigs* pelo montador *ABYSS*. Dois outros montadores foram utilizados: *SPAdes* e *Velvet*, os quais produziram-se 95.736 e 1.735.434 *contigs*, respectivamente. Estes *contigs* apresentaram *hits* com vírus de planta a partir do algoritmo *tblastx* utilizando o RefSeq viral. Os *hits* foram em sua maioria referentes a sequências de espécies das famílias virais *Caulimoviridae*, *Tombusviridae*, *Rhabdoviridae* e *Potyviridae*. Entre estes, seis genomas foram montados *in silico* completo ou parcialmente a partir das sequências de genoma de referência, que são *Sonchus yellow net nucleorhabdovirus* (SYNV), *Potato yellow dwarf nucleorhabdovirus* (PYDV), *Alfalfa dwarf cytorhabdovirus* (ADV), *Jasminum ringspot virus* (JRSV), *Carrot mottle mimic virus* (CMoMV) e *Rosa rugosa leaf distortion virus* (RrLDV). Após comparação com os genomas de referência, todas as espécies apresentaram indicativo de novas espécies para ciência, seguindo os critérios taxonômicos pré-estabelecidos para cada grupo. Para a identificação viral das espécies acima listadas em cada amostra de planta ornamental coletada e no *pool*, fez-se detecção

via RT-PCR a partir de *primers* confeccionados com base na sequência obtida via NGS. *Amplicons* foram obtidos para as espécies virais SYNV e CMoMV em *Coreopsis lanceolata*; para PYDV em *Pachystachys lutea* e RrLDV em *Jasminum nitidum*. Nenhum *amplicon* foi obtido quando se utilizou *primers* para ADV e JRSV nas amostras individuais. Duas prováveis espécies virais novas foram selecionadas para estudos posteriores de recuperação do genoma completo e das extremidades 5' e 3'. As sequências destas duas espécies, provenientes de *Coreopsis lanceolata* e *Pachystachys lutea*, respectivamente, apresentaram nesta ordem, similaridade com SYNV e PYDV e foram denominadas *Coreopsis mottle-associated virus* (CMAV) e *Pachystachys mosaic-associated virus* (PMAV). As sequências completas do genoma viral de CMAV e PMAV foram recuperadas via método de Sanger. Análises filogenéticas indicaram que CMAV apresentou 56,99% de identidade nucleotídica com o genoma completo de SYNV. O genoma de PMAV apresentou relação com PYDV, com identidade nucleotídica de 51,88% para o genoma completo. Ambos os vírus estão agrupados em um clado monofilético do gênero *Nucleorhabdovirus*, família *Rhabdoviridae*. Acredita-se, portanto, que os vírus CMAV e PMAV representam espécies novas do gênero *Nucleorhabdovirus*.

Palavras-chave: Bioinformática, *Coreopsis mottle-associated virus*, Filogenia, *Nucleorhabdovirus*, *Pachystachys mosaic-associated virus*, *Rhabdoviridae*, Viroma.

1. INTRODUÇÃO

As plantas ornamentais utilizadas para paisagismo em áreas públicas do Distrito Federal são produzidas no Viveiro I da NOVACAP. Esse viveiro é responsável pela produção e distribuição de todas as flores e plantas ornamentais plantadas em solo público no DF (NOVACAP, 2017). Brasília é uma cidade tipicamente arborizada e dotada de canteiros com flores. Essa característica da cidade, fixada já como cultura, é um dos diversos cartões postais que a mesma apresenta. Além disso, a produção de plantas ornamentais no DF é a 10ª no país, gerando, em 2014, um total de R\$ 125.072.000 em valor de mercado. Assim, o estudo de patógenos acometendo essas plantas em sua origem, ou seja, no local de produção, mostra-se de profundo interesse tanto do produtor, como da agricultura em geral, já que as plantas ornamentais difundidas por todo o DF podem atuar como fontes de inóculo para doenças que acometem grandes culturas (Alexandre *et al.*, 2010).

Vírus, pela sua natureza diminuta e capacidade de rápida dispersão, seja por unidades propagativas de material vegetal, por vetores e/ou pela ação humana, são de interesse para a definição de procedimentos fitossanitários envolvendo plantas ornamentais. Como a produção destas envolve, muitas vezes, técnicas de corte (como em flores de corte, multiplicação de plantas via estaquia, entre outros), as plantas ficam propensas a aquisição e posteriormente propagação de patógenos, já que muitos podem se comportar de forma latente. Além disso, sabe-se que muitas das entradas de patógenos no país ocorre por importação de material propagativo (estacas, por exemplo) infectado (Alexandre *et al.*, 2010).

O mercado de plantas ornamentais, atualmente em ampla expansão não só no Brasil como no mundo, é caracterizado por um fluxo grande e contínuo de entrada e saída de diferentes plantas. Acredita-se em uma tendência que as plantas ornamentais, caso não sejam estudadas continuamente quanto à detecção viral, passem a portar diferentes vírus potencialmente destrutivos à própria cultura e produção de ornamentais e a outras culturas (Harakava *et al.*, 2008).

Como informações sobre a distinção entre os vírus em nível de espécie ou gênero ainda são escassas em se tratando de plantas ornamentais, pesquisas envolvendo esse tipo de patógeno podem ser fundamentais para contribuir para o bom desempenho na produção e, portanto, para o mercado em geral. Com o advento de novas tecnologias de sequenciamento, como o de NGS, pode-se adotar essas novas metodologias para um

estudo de diagnose mais amplo. Sabendo-se que ainda pouco se sabe de vírus em geral, pesquisas envolvendo amostras pouco trabalhadas pode ser de alto interesse para a virologia vegetal (Mokili *et al.*, 2012). Com o NGS, pode-se obter diferentes sequências virais mesmo sem referência ou quando não há indício prévio de doença ou agente causal. É uma vantagem significativa para plantas que são pouco ou não estudadas, já que nestas não há ou pouco são os relatos de doenças virais (Radford *et al.*, 2012); o histórico de doenças destas seria escasso não havendo, portanto, numerosos vírus de referência. Assim, o NGS, aliado à metagenômica, pode gerar novas descobertas de espécies de vírus ainda não relatadas em plantas ornamentais ou para a ciência.

A identificação de novos vírus por meio de NGS segue um rigoroso fluxo de trabalho envolvendo ferramentas de bioinformática. A bioinformática é um novo ramo que une Biologia e Ciência da Computação que cresceu com a ascensão das novas tecnologias da genômica. Lidar com os milhões de dados gerados pelas plataformas de sequenciamento demanda tempo, cuidado e trabalho (Radford *et al.*, 2012). O processamento de dados de sequenciamento pode ser oneroso e problemático, demandando o uso de diferentes programas e a sua correta configuração (Scholz *et al.*, 2012). A elaboração de um genoma “virtual” (*in silico*) demanda, por enquanto, que o mesmo seja amplificado e sequenciado por outros métodos, como Sanger. Isto é, deve-se confirmar o genoma do suposto vírus via métodos tradicionais moleculares, como PCR e *Rapid Amplification cDNA Ends* (RACE). A recuperação demanda sucessivas ampliações de diferentes regiões da sequência obtida e da região terminal 3' e 5'. Para estas últimas, a recuperação também completa as sequências, utilizando-se o protocolo RACE, que consiste na amplificação das regiões terminais de uma sequência a partir de um segmento pré-determinado, por meio de *primers* específicos a sequência da amostra e ao segmento (Matz *et al.*, 1999).

Plantas ornamentais produzidas no Viveiro I da NOVACAP apresentavam sintomas típicos de infecção viral, como mosaico, mosqueado, encarquilhamento e outros. Em plantas ornamentais, sintomas em folhagens e inflorescências podem comprometer a estética da planta, inviabilizando a comercialização e distribuição da mesma. O objetivo deste trabalho foi buscar e identificar espécies virais a partir de dez plantas ornamentais que apresentavam sintomas típicos de vírus produzidas no Viveiro I da NOVACAP via NGS, bioinformática e ensaios moleculares.

2. MATERIAL E MÉTODOS

2.1. Coleta e obtenção das amostras de plantas ornamentais a serem analisadas

Plantas ornamentais com sintomas típicos daqueles causados por espécies virais, como mosaico, mosqueado, amarelecimento, clorose, anéis cloróticos, encarquilhamento e epinastia, foram selecionadas e coletadas no Viveiro I da NOVACAP nos dias 10/11/2015 e 27/11/2015. Foram, no total, folhas de dez amostras de plantas ornamentais obtidas, todas sintomáticas e coletadas em pontos ao longo do viveiro, como em telados, estufas e a céu aberto (**Tabela 1**). A seleção foi realizada com base nos sintomas típicos de vírus observados no momento da coleta. Todas as plantas foram fotografadas *in situ* (**Figura 1**).

Tabela 1. Plantas ornamentais (nome comum e científico) coletadas para estudo e com base nos sintomas observados no momento da coleta no Viveiro I da NOVACAP (Companhia Urbanizadora da Nova Capital do Brasil).

Família botânica	Nome popular	Nome científico	Sintoma associado
Acanthaceae	Camarão amarelo	<i>Pachystachys lutea</i>	Mosaico
Araceae	Antúrio	<i>Anthurium lindmanianum</i>	Manchas cloróticas
Araceae	Jibóia	<i>Epipremnum aureum</i>	Manchas cloróticas
Arecaceae	Palmeira Pinanga	<i>Pinanga kuhlii</i>	Mosaico
Asteraceae	Margaridinha-amarela	<i>Coreopsis lanceolata</i>	Mosqueado
Balsaminaceae	Beijinho-pintado	<i>Impatiens hawkeri</i>	Ápice encarquilhado
Geraniaceae	Gerânio	<i>Pelargonium</i> sp.	Anéis verdes/amarelos
Iridaceae	Lírio-da-praia	<i>Neomarica candida</i>	Mosqueado
Oleaceae	Jasmim-estrela	<i>Jasminum nitidum</i>	Anéis cloróticos
Solanaceae	Marianinha	<i>Streptosolen jasmonii</i>	Amarelecimento

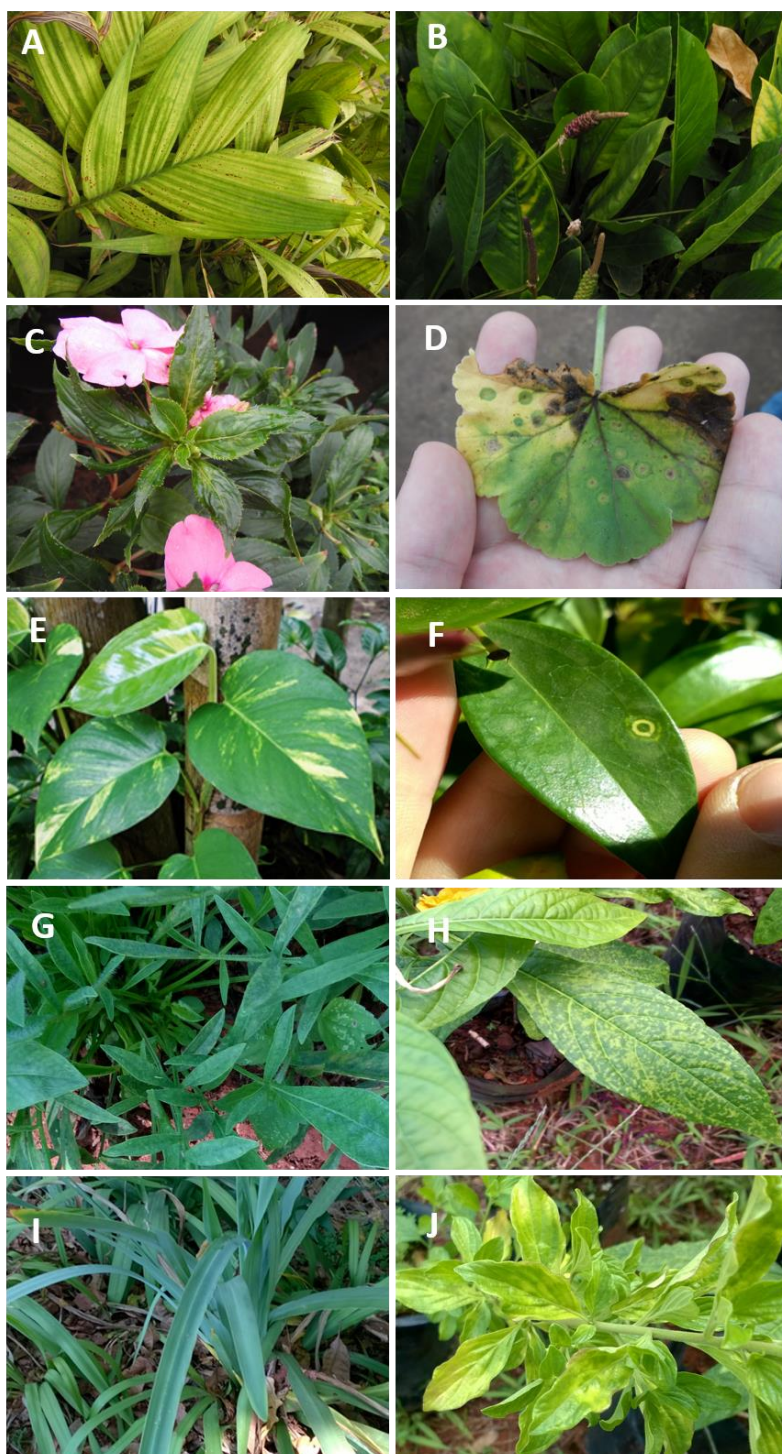


Figura 1. Sintomas similares àqueles induzidos por vírus de plantas e observados nas dez plantas ornamentais selecionadas e coletadas no Viveiro 1 da NOVACAP (Companhia Urbanizadora da Nova Capital do Brasil). As amostras são: A – *Pinanga kuhlii* (palmeira pinanga; com sintoma de mosqueado); B – *Anthurium lindmanianum* (antúrio; com sintoma de amarelecimento e clorose); C – *Impatiens hawkeri* (beijinho; com sintoma de ápice encarquilhado); D – *Pelargonium* sp. (gerânio; com sintoma de anéis cloróticos verdes e amarelos); E – *Epipremnum aureum* (jibóia; com sintoma de amarelecimento); F – *Jasminum nitidum* (jasmim-estrela; com sintoma de anéis cloróticos); G – *Coreopsis lanceolata* (margaridinha; com sintoma de mosqueado); H – *Pachystachys lutea* (camarão-amarelo; com sintoma de mosaico); I – *Neomarica candida* (Lírio-da-praia; com sintoma de amarelecimento); J – *Streptosolen jasmonii* (marianinha; com sintoma de amarelecimento e mosqueado).

2.2. Enriquecimento viral, extração de RNA e preparo de um *pool* de amostras para Next Generation Sequencing - NGS

As amostras obtidas no viveiro foram imediatamente levadas ao Departamento de Biologia Celular do Instituto de Biologia (CEL-IB) da Universidade de Brasília (UnB), onde foram processadas e armazenadas a -80° C. Cada espécie de planta ornamental foi armazenada separadamente.

As dez amostras selecionadas foram submetidas ao processo de enriquecimento de partículas virais mediante semi-purificação, seguido de extração de RNA. Para isso, 2 gramas de folhas de cada planta ornamental coletada foram agrupadas totalizando 20 g. As folhas foram trituradas na presença de nitrogênio líquido, adicionando-se, posteriormente, tampão fosfato de potássio (pH 8,0) 0,1 M, EDTA 1 mM e β -mercaptoetanol (0,2%). Posteriormente, o *pool* foi filtrado com gaze previamente esterilizado e centrifugado por 20 minutos (min) a 5.000 rotações por minuto (rpm) (3.836 x g) em uma centrífuga preparativa à temperatura de 4°C. O sobrenadante foi então transferido para 2 tubos de ultracentrífuga e adicionou-se uma solução de sacarose 20% (20 mL) para a formação do colchão de sacarose e centrifugado a 33.000 x g por 2 horas (h) à vácuo e à temperatura de 4°C. O RNA total foi extraído do precipitado com o kit TRIzol RNA Isolation Reagent (ThermoFisher Scientific) seguindo o manual do fabricante. Finalmente, o RNA total foi quantificado por espectrofotometria (Nanodrop) e mantido em *RNA Stable* (Sigma-Aldrich) para então ser enviado para o sequenciamento na MacroGen Inc. (Coréia do Sul), utilizando a tecnologia da plataforma *Illumina HiSeq 2000*. Para o sequenciamento propriamente dito, foi utilizado o *TruSeq SBS Kit v3* como reagente, *TruSeq Stranded Total RNA LT Sample Prep Kit* para preparo da biblioteca e o *HCS v2.2* como *software* de controle de sequenciamento. O fluxograma do procedimento de sequenciamento envolveu o preparo da amostra, produção da biblioteca, agrupamento e amplificação dos *reads*, sequenciamento dos *reads* e análise dos dados.

2.3. Análise computacional das sequências obtidas via Next Generation Sequencing - NGS

As sequências obtidas pelo NGS foram analisadas, montadas e interpretadas nos computadores de alto desempenho do CEL-IB da UnB. As análises e interpretações visuais das sequências foram realizadas por meio do programa *Geneious R9* (Biomatters Ltd) (Kearse *et al.*, 2012).

Os arquivos brutos de sequenciamento, contendo as milhões de pequenas sequências (*reads*), foram armazenados e salvos nos computadores de trabalho e nos servidores *Galaxy* (usegalaxy.org – The Galaxy Project). Os *reads* foram utilizados para montagem da sequência consenso (*contigs*). Para isso, as sequências foram submetidas a análise de controle de qualidade via programa *FastQC* (Babraham Bioinformatics). As regiões consideradas de baixa qualidade de cada *read* foram eliminadas por meio do software *Trimmomatic* (Bolger *et al.*, 2014). As sequências resultantes foram, em seguida, utilizadas para a montagem da sequência consenso em montadores (do inglês, *assemblers*) de maneira a se obter os *contigs*. Foram três *softwares* de montadores utilizados: *Velvet* (Zerbino and Birney 2008), *SPAdes* (Simpson, Wong *et al.* 2009) e *ABySS* (Zerbino & Birney, 2008, Simpson *et al.*, 2009, Bankevich *et al.*, 2012). Cada grupo de *contigs* gerado pelos diferentes montadores foi individualmente submetido à análise pelo algoritmo *tblastx* (<https://blast.ncbi.nlm.nih.gov/Blast>) contra o banco de dados viral *RefSeq* (*GenBank*), este montado contendo todas as sequências virais de referência do *GenBank* até a data de setembro, 2016, excetuando-se sequências de bacteriófagos, totalizando 5.495 sequências.

2.4. Seleção e montagem do genoma das duas espécies virais: Cariopsis mottle-associated virus (CMaV) e Pachystachys mosaic-associated virus (PMaV) para caracterização molecular

Os *contigs* submetidos ao alinhamento local pelo *tblastx* que resultaram em um *hit* com sequências de vírus de plantas foram selecionados. Entre esses, uma nova seleção foi feita com base nas condições dos *contigs*, com estes superando 1000 pb em tamanho e/ou com mais de 100 *contigs* em número, de modo a identificar possíveis vírus que estejam em maior número nas amostras. Estas sequências selecionadas, então, foram agrupadas com base na espécie viral de maior identidade detectada pelo *tblastx*. Uma vez analisado, cada grupo de *contigs* foi montado *de novo* a partir dos *reads* de maneira a se obter, computacionalmente, o genoma parcial ou completo de cada sequência viral. A montagem ocorreu utilizando os *reads* como referência, até se obter uma sequência viral de tamanho próximo à espécie de relação pelo *tblastx*. A análise visual e comparativa, *tblastx* e a montagem dos genomas foram feitos no programa *Geneious R9* (Kearse *et al.*, 2012).

2.5. Detecção de seis espécies virais por meio de RT-PCR (*Reverse Transcriptase – Polymerase Chain Reaction*) nas plantas ornamentais selecionadas

Para a detecção das espécies virais obtidas via sequenciamento NGS nas amostras coletadas, foram desenhados e sintetizados *primers* específicos para cada uma das seis espécies virais selecionadas. Objetivou-se posicionar/localizar os *primers* para amplificação de *amplicons* de 500 a 600 pb, dentro de genes conservados quando em comparação com suas respectivas sequências de referência. Os seis vírus foram detectados a partir da amplificação de parte do gene que codifica a polimerase, por essa ORF estar presente em todas as sequências montadas após NGS, incluindo aquelas para as quais não se obteve o genoma completo. Todos os *primers* possuem 20 pb e a temperatura de anelamento (T_a) seguiu sempre a temperatura de *melting* (T_m) – 2° C. Além disso, todos os *primers* sintetizados nesse trabalho foram desenhados pelo *software Geneious R9 (Tabela 2)*.

Inicialmente, a detecção de cada vírus foi realizada utilizando *primers* desenhados com base nas sequências virais montadas a partir do sequenciamento NGS de RNA total do *pool* de amostras (o mesmo utilizado no sequenciamento NGS). Os resultados para cada par de *primers* foram, então, usados como controle positivo para os ensaios seguintes. Posteriormente, testou-se os mesmos *primers* nas amostras de maneira individual, a fim de identificar a hospedeira para cada sequência viral.

Tabela 2. *Primers* usados para detecção de seis espécies virais no *pool* e nas amostras individuais de plantas ornamentais.

<i>Primers*</i>	Sequência (5'-3')	Tamanho do amplicon (pb)**	Temperatura de melting (Tm) (°C)	GC***(%)
ADV1-F	CGAATCAGTTGAAAGAGGCCG	638	57.2	50
ADV2-R	CTTGATTATCACCCATCCCC	638	55.2	50
PMaV-F	GTGACTTCCTAAGATTGGCC	611	55.8	50
PMaV-R	GTATGTCATCTATGGACGCG	611	55.9	50
CMaV-F	CATCGCTATCTGACTTCCGG	545	57.9	55
CMaV-R	CTCTCTTGACATCTTAGCCC	545	55.0	50
JRSV1-F	AGTGACTTGGTAATGAGGGG	532	56.5	50
JRSV2-R	CCAAGTGATGTTTTGTCGGC	532	58.0	50
RrLDV1-F	CTTGAGTCAGAGCTAATCCG	569	55.4	50
RrLDV2-R	CGACCAACATGTTTCTGTGG	569	57.3	50
CMoMV-F	CTAGTGAGATCATGCAGGCC	625	57.8	55
CMoMV-R	AAGAGAGGAACATGAAGCCG	625	57.6	50

* F: *Forward*, R: *Reverse*. ADV: *Alfalfa dwarf virus*; PMaV: *Pachystachys mosaic-associated virus*; CMaV: *Coreopsis mottle-associated virus*; JRSV: *Jasminum ringspot virus*; RrLDV: *Rosa rugosa leaf distortion virus*; CMoMV: *Carrot mottle mimic virus*. **pb: pares de bases e ***GC: Guanina e citosina.

Para o procedimento de detecção propriamente dito, reações de *Reverse Transcriptase – Polymerase Chain Reaction* (RT-PCR) foram realizadas. Nessa metodologia, produz-se cDNA a partir do RNA total extraído da amostra. O cDNA, tanto para o *pool* como para as amostras individuais, foi sintetizado utilizando a transcriptase reversa do kit SuperScript III (Invitrogen), a partir do *primer* reverso específico para cada sequência viral (50 μ M). A reação de transcrição reversa foi realizada utilizando 5 μ L de RNA total (aproximadamente 200 ng/ μ L), 1 μ L de dNTP (10 mM), 1 μ L do *primer* reverso e 5 μ L de H₂O, seguida de 5 min de incubação a temperatura de 65 °C para abrir a fita de RNA. Posteriormente, foram adicionados 4 μ L de 5X First-Strand Buffer, 2 μ L de DTT (0,1 M), 1 μ L da enzima SuperScript III (200 U/ μ L) e 1 μ L de RNase out (Thermofisher Scientific) (40 U/ μ L). A temperatura de incubação utilizada nesta etapa foi de 55 °C, seguida por um período de incubação de 70 °C por 15 min para a inativação da enzima.

A LongAmp DNA Polimerase (New England Biolabs) foi utilizada para a realização da PCR, seguindo instruções do fabricante. Para a PCR, utilizou-se 1 μ L dos *primers forward* e *reverse* específicos para cada sequência viral (10 μ M), 5 μ L de 5X LongAmp Reaction Buffer, seguida de 1 μ L de dNTP, 1 μ L de LongAmp Taq DNA Polymerase (2,5 U/ μ L), 1 μ L do cDNA e 15,5 μ L de H₂O, para uma reação final de 25 μ L. As condições do termociclador foram: “*hot-start*”, por 1 minuto, a 80° C; desnaturação inicial, por 2 minutos, a 94° C seguida de 35 ciclos de desnaturação por 30 segundos a 94° C; anelamento (conforme **Tabela 2**) por 40 segundos, extensão por 40 segundos a 65°C e extensão final por 10 minutos a 65°C.

2.6. Estratégias para recuperação do genoma completo de *Coreopsis mottle-associated virus* (CMAV) e *Pachystachys mosaic-associated virus* (PMAV)

Entre as seis sequências de prováveis espécies virais novas, duas foram escolhidas para serem completamente sequenciadas via método convencional Sanger, segundo o critério: genoma montado completo *in silico*, número de *reads* por mapeamento acima de 85.000 e sequências detectadas em amostras individuais.

As sequências de referência correspondem a duas espécies da família *Rhabdoviridae*: *Sonchus yellow net nucleorhabdovirus* - SYNIV e *Potato yellow dwarf nucleorhabdovirus* - PYDV. De acordo com critérios de nomenclatura, que serão melhor abordados e discutidos em resultados e discussão, o caráter hospedeira foi considerado e as duas prováveis espécies novas provenientes de *Cariopsis lanceolata* (marianinha) e

Pachystachys lutea (camarão amarelo) passaram a ser denominadas, então, de Cariopsis mottle-associated virus (CMAV) e Pachystachys mosaic-associated virus (PMAV), respectivamente.

Inicialmente realizou-se extração de RNA total das amostras de marianinha e camarão amarelo utilizando o protocolo de TRIzol (ThermoFisher Scientific) de acordo com manual do fabricante. A qualidade e concentração de RNA foram analisadas em gel de agarose (RNase free) a 1%. Reações para síntese de DNA complementar (cDNAs) foram realizadas conforme descrição no item 2.5. Após este procedimento, reações de PCR para cada região do genoma foram realizadas com a enzima LongAmp DNA Polimerase (ThermoFisher Scientific), com as mesmas condições descritas para o ensaio de detecção (item 2.5), porém com as seguintes especificações no termociclador: “*hot-start*”, por 1 minuto, a 80° C; desnaturação inicial, por 2 minutos, a 94° C; 35 ciclos de desnaturação por 30 segundos a 94° C; anelamento por 40 segundos a 55° C, extensão por 150 segundos a 65°C e extensão final por 10 minutos a 65°C.

Como estratégia para recuperação do genoma completo das espécies aqui denominadas CMAV e PMAV o genoma foi dividido em cinco (1, 2, 3, 4 e 5) e quatro (1, 2, 3 e 4) regiões respectivamente, de modo que cada um delas apresentasse em torno de 2900 pb (**Tabela 3**). As sequências obtidas via NGS foram usadas para desenho e disposição dos *primers* de modo que cada *amplicon* com tamanho em torno de 2900 pb apresentasse uma sobreposição de, aproximadamente 200 pb em relação ao outro, como ilustrado na **Figura 2**.

Tabela 3. Sequências dos *primers* utilizados em cada região genômica para as sequências de Coreopsis mottle-associated virus (CMaV) e Pachystachys mosaic-associated virus (PMaV).

Primers F e R *	Sequência (5'-3')	Tm**(°C)	GC ***(%)	Região do genoma
CMaV_Ctg1F	ATGAGCACACCACACCAGCAAT	57.4	50.0	1
CMaV_Ctg1R	TCTCTTCCCATTGTCTGTGCT	56.0	47.6	1
CMaV_Ctg2F	GGGACTCCGTATCTTGTTCTTG	55.1	50.0	2
CMaV_Ctg2R	AAATGATTGCCCATACAGGACT	54.4	40.9	2
CMaV_Ctg3F	AGTAGCAGGTTCGTCTTGCTC	56.6	52.4	3
CMaV_Ctg3R	AGGATCGTCAAGGTTGGTCTTAT	55.7	43.5	3
CMaV_Ctg4F	GGTGTGGTGTGTTGTGCGTATA	55.2	50.0	4
CMaV_Ctg4R	CCTAATGGATGTGGTGTGTTGTGAC	55.9	47.8	4
CMaV_Ctg5F	CAATCCACCACAATCAACAAAATAGC	55.3	38.5	5
CMaV_Ctg5R	CCTCCCAGATGGTCCTTAATC	54.5	52.4	5
PMaV_Ctg1F	ACTCAAAGCAGGGTCCAG	54.6	55.6	1
PMaV_Ctg1R	CAT CCT GGA AGA CAG GGC	55.4	61.1	1
PMaV_Ctg2F	AGC ACA TGT CCG AAG GAG	54.6	55.6	2
PMaV_Ctg2R	CCC CTG ACA CGG TTA TCA TC	55.1	55.0	2
PMaV_Ctg3F	GGTCTTGCCGCTTACACA	55.4	55.6	3
PMaV_Ctg3R	ATA ACA GGC TAT TGG CCC	52.0	50.0	3
PMaV_Ctg4F	TGC AGT GCT CAG AGA ACG	55.0	55.6	4
PMaV_Ctg4R	ATC CGC TTC ACA ATG TCC G	55.5	52.6	4

* F: *Forward*, R: *Reverse*. Todas as combinações geraram *amplicons* de aproximadamente 2.900 pares de bases. **: Temperatura de melting e***GC: Conteúdo de guanina e citosina.

Para a confirmação das sequências, cada produto da amplificação por PCR foi eluído do gel de agarose a 1% (*GFX PCR DNA and Gel Band Purification* - GE Healthcare), seguindo as recomendações do fabricante. Após a eluição, os fragmentos de PCR foram submetidos ao sequenciamento convencional via Sanger direto, realizado pela Macrogen Inc. (Seoul, Coréia). Cada sequência resultante do método Sanger foi avaliada, com as regiões de baixa qualidade removidas e montadas com base na sobreposição entre sequências, pelo *software Geneious R9*. Reações adicionais por *primer walking* foram realizadas, quando necessário, para a completa determinação da sequência do produto de PCR.

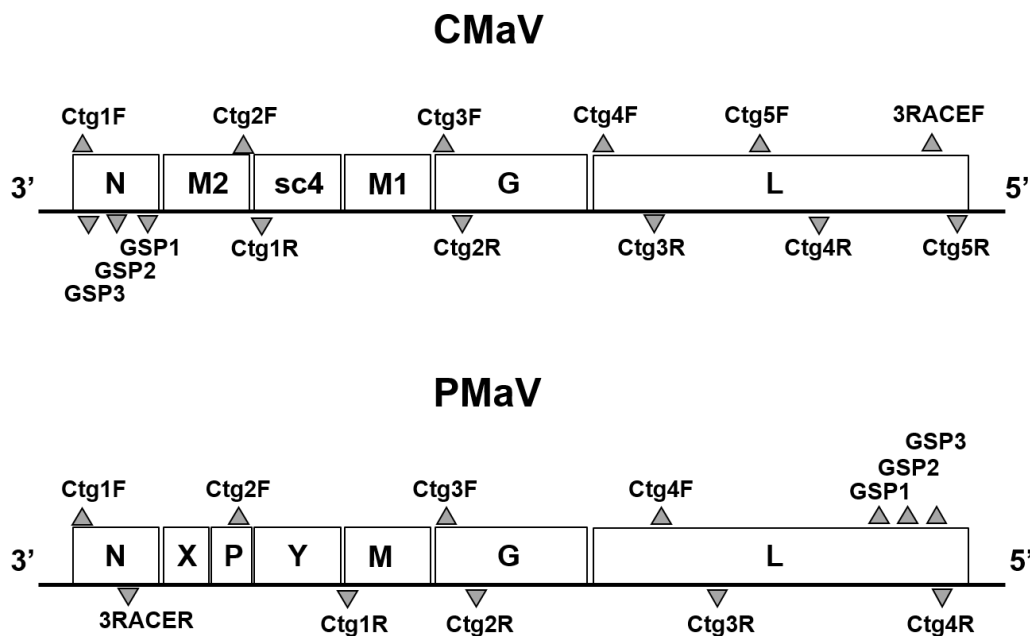


Figura 2. Disposição dos *primers* específicos para as duas sequências virais: Coreopsis mottle-associated virus (CMAV) e Pachystachys mosaic-associated virus (PMAV), desenhadas a partir das sequências obtidas via *Next Generation Sequencing* – NGS, ao longo do respectivo genoma. Os retângulos representam os genes, sendo que os mesmos foram supostos por analogia aos vírus de referência (*Sonchus yellow net nucleorhabdovirus* - SYNV e *Potato yellow dwarf nucleorhabdovirus* - PYDV, respectivamente). As setas representam a localização próxima de cada *primer* no genoma. Os possíveis genes são referentes às seguintes proteínas: N – nucleocapsídeo; M – proteína de matriz; G – glicoproteína; L – polimerase; sc4 – possível supressor de silenciamento gênico/proteína de movimento; P – fosfoproteína; X – proteína “X”; Y – possível proteína de movimento. F: *Forward*; R: *Reverse*. Ctg: “*Contig*” equivalente a parte do genoma dividido. GSP: *primers* internos do 5’ RACE (*Rapid Amplification of cDNA Ends*).

2.7. Recuperação das regiões 3’ e 5’ dos genomas de *Cariopsis mottle-associated virus* (CMAV) e *Pachystachys mosaic-associated virus* (PMAV) por meio de RACE (*Rapid Amplification of cDNA Ends*)

O protocolo RACE consiste na amplificação por RT-PCR das extremidades 3’ e 5’ de um mRNA (RNA mensageiro) a partir de um segmento interno de sequência pré-

determinada. No 3'RACE, utiliza-se uma cauda poli(A), existente ou não na extremidade 3' do mRNAs alvo, como sítio de anelamento para um *primer* genérico, a fim de, em conjunto com um *primer* específico interno, amplificar a sequência compreendida entre ambos. Na falta da cauda, adiciona-se a mesma por meio da poli-A Polimerase (New England Biolabs) (Matz *et al.*, 1999).

O 5'RACE consiste em diferentes ampliações por RT-PCR para isolar a extremidade 5' não conhecida de um RNA a partir de fragmentos internos de sequências conhecidas. A síntese de cDNA é feita utilizando um *primer* para uma região conhecida (sendo este o mais “afastado” da região 5'). Uma sequência de Citidina (Citosina) é adicionada a extremidade 3' da fita nascente com a enzima Terminal Desoxinucleotídeo Transferase (TdT) (New England Biolabs). A primeira PCR é realizada com um segundo *primer* específico, mais interno em relação ao primeiro e um *primer* complementar a região homopolimérica adicionada, que possui uma região estratégica (âncora) que vai ser utilizada na próxima reação de PCR. Na segunda etapa, um terceiro *primer* específico, mais interno dos três, é utilizado, juntamente com o *primer* que anela especificamente à região âncora da primeira PCR (Matz *et al.*, 1999).

Para o 3'RACE e PCR, foi necessário adicionar cauda poli(A), por meio da poli(A) polimerase (New England Biolabs). Para a reação de adição de cauda poli-A, 10 µL de RNA total (aproximadamente 200 ng/µL) foram utilizados juntamente com 5 µL de H₂O, 2 µL de tampão 10x Poli-A (New England Biolabs), 2 µL de ATP (10 mM) e 1 µL de Poli(A) Polimerase (5 U/µL). A reação foi incubada durante 30 minutos a 37° C e, para inativação da enzima, a solução foi incubada a 65° C por 20 minutos. O cDNA foi sintetizado utilizando a transcriptase reversa SuperScript III (Invitrogen) (200 U/µL), utilizando os *primers* descritos na **Tabela 3**, cujas condições foram especificadas no item 2.5. A reação de transcrição reversa consistiu de 5 µL de RNA total (amostra poliadenilada), 1 µL de dNTP 10 mM, 1 µL do *primer* Oligod50TM4 (Filho *et al.*, 2013) (50 µM) e 5 µL de H₂O, seguida de 5 min de incubação a temperatura de 65 °C. Posteriormente foi adicionado 4 µL de 5X First-Strand Buffer, 2 µL de DTT (0,1 M), 1 µL da enzima SuperScript III e 1 µL de RNase out (ThermoFisher Scientific) nas mesmas concentrações descritas em 2.5. A temperatura de incubação utilizada nesta etapa foi de 50 °C por 50 minutos, seguida por um período de incubação de 85 °C por 5 minutos para a inativação da enzima. Para a amplificação da região 3', utilizou-se um protocolo típico para a PCR com a enzima LongAmp Taq DNA Polimerase, dessa vez utilizando o *primer* M4 e um específico a região. Utilizou-se, portanto, 1 µL dos *primers* M4 (Chen *et al.*,

2001) e específico (10 μM), 5 μL de 5X LongAmp Reaction Buffer, seguida de 1 μL de dNTP 10 mM, 1 μL de LongAmp *Taq* DNA Polymerase (2,5 U/ μL) e 1 μL do cDNA e 15,5 μL de H_2O , para uma reação de final de 25 μL . As condições do termociclador foram: “*hot-start*”, por 1 minuto, a 80° C; desnaturação inicial por 2 minutos a 94° C; seguida de 35 ciclos de desnaturação por 40 segundos a 94° C; anelamento por 60 segundos a 56° C; extensão por 90 segundos a 65° C e extensão final por 10 minutos a 65° C.

Para amplificar a região da extremidade 5', o cDNA foi sintetizado a partir da enzima SuperScript III Reverse Transcriptase. Para isto, 10 μL de amostra (RNA; 200 ng/ μL) foram utilizados, seguido de 1 μL de dNTPs (10 mM) e 1 μL do *primer* mais afastado da extremidade 5' (50 μM). A amostra foi incubada a 75° C por 5 minutos e, logo em seguida, foram adicionados os seguintes reagentes: 4 μL do tampão 5x da enzima, 2 μL DTT (0,1 M), 1 μL de RNase out (40 U/ μL), 1 μL da enzima SuperScript III (200 U/ μL). Incubou-se a 52° C por 1 hora e, em seguida, a 75° C por 20 minutos. Imediatamente após, adicionou-se os reagentes finais, 1 μL de RNase H (5 U/ μL) (New England Biolabs) e 0,1 μL de RNase A (ThermoFisher Scientific) (10 mg/ μL) para uma última incubação a 37° C por 30 minutos. Uma vez com o cDNA, a amostra foi purificada utilizando-se kit de purificação da GE Healthcare (*illustra GFX PCR DNA and Gel Band Purification Kit*), seguindo o protocolo do fabricante. Após purificação, adiciona-se a cauda poli-C a partir da enzima Terminal Transferase. Assim, adicionou-se 2 μL de tampão da enzima, 2 μL de cloreto de cobalto (CoCl_2) 2,5 mM, 0,5 μL de dCTP (5 mM) e 0,2 μL da enzima TdT (New England Biolabs) (20 U/ μL). Completou-se o restante com água (15,3 μL). O tubo foi incubado a 37° C por 1 hora e, então, a 70° C por 10 minutos. Com o cDNA purificado e somada de cauda poli-C, realizou-se duas PCR. A primeira foi idêntica a utilizada no 3'RACE, apenas com os *primers* específicos “5R-Âncora” e “GSP2” – mais interno que o primeiro (**Tabela 4**). A segunda PCR, por sua vez, utiliza o produto da reação da primeira como cDNA. Para isso, usou-se apenas 0,1 μL de cDNA. Os *primers*, desta vez, foram o “5' Nested” e o “GSP3” – mais interno que o segundo (**Tabela 4**). As condições restantes foram idênticas.

Para a determinação da região 3' e 5' UTR, utilizou-se o mesmo RNA total extraído e utilizado nos demais ensaios.

Tabela 4. Primers utilizados nos protocolos 3' e 5' RACE (*Rapid Amplification of cDNA Ends*) para determinação de sequências de extremidades das espécies Coreopsis mottle-associated virus (CMaV) e Pachystachys mosaic-associated virus (PMaV).

Primers F e R*	Sequência (5'-3')	Tm** (°C)	GC*** (%)	Região UTR****
Oligod50TM4 ¹	GTTTTCCCAGTCACGACTTAATTAA(T) ₅₀	63.5	12.0	3'
3RACEF CMaV	GCCAAACATTAGTTGACACTCTC	54.0	43.5	3'
3RACEF PMaV	GTGGTCTCGGAGATTGTGC	55.8	57.9	3'
M4 ²	GTTTTCCCAGTCACGACT	50.6	52.9	3'
GSP1 CMaV	TGAGGAACTACCTGTGTCCG	56.1	55.0	5'
GSP1 PMaV	GCATTGAGAACCAGAGAAGGC	56.1	52.4	5'
5R-Âncora	AAACCACGACACCTCCAAGCAAGGGGIIGG GIIGGGIIG	77.0	56.4	5'
GSP2 CMaV	CGGTGGAATAATCAGCACATAG	53.4	45.5	5'
GSP2 PMaV	GCGAGGAGGTTTCCTCAG	55.0	64.7	5'
5' Nested	ACCACGACACCTCCAAGCAAG	59.8	57.1	5'
GSP3 CMaV	ATGTCACTCTCCGTCATGTG	54.4	50.0	5'
GSP3 PMaV	ATCCAGCCTGTGCGAGCAT	56.5	55.6	5'

* F: *Forward*, R: *Reverse*. **: Temperatura de melting e***GC: Guanina e citosina. **** *untranslated region* (região não traduzível). 1: (Filho *et al.*, 2013) 2: (Chen *et al.*, 2001)

O protocolo RACE de CMaV foi feito com base na fita antiviral (senso positivo), enquanto que para PMaV, o oposto, ou seja, a fita viral (senso negativo). Isso inverteu os polos dos *primers* entre as duas espécies.

2.8. Análises Filogenéticas e comparação de sequências para o *Coreopsis mottle-associated virus* (CMaV) e *Pachystachys mosaic-associated virus* (PMaV)

As análises filogenéticas e de comparação de sequências das amostras foram feitas com base nos alinhamentos globais múltiplos (MSA) e/ou *pairwise*, utilizando o algoritmo do MAFFT (*Multiple Alignment using Fast Fourier Transform* – Kazutaka Katoh). Os alinhamentos *pairwise* utilizaram sequências de aminoácidos e nucleotídeos obtidos por banco de dados (GenBank/EMBL/DDBJ) (**Tabelas 5**).

Após o alinhamento, a análise evolutiva foi realizada por árvores filogenéticas construídas com base nos alinhamentos *pairwise* entre sequências. O método (algoritmo) utilizado foi o *Maximum Likelihood*, pelo seu código altamente sofisticado que permite um modelo evolucionário mais realista ao se analisar todas as estruturas possíveis da árvore. Os demais parâmetros foram alterados até se obter *bootstraps values* maiores. O *software* utilizado na construção da árvore foi o MEGA 6 (Tamura *et al.*, 2013). Para a construção da matriz de identidade, foi utilizado o software SDT v1.2 (*Sequence Demarcation Tool*) (Muhire *et al.*, 2014). Para a análise e construção das matrizes e árvores filogenéticas, foram utilizadas sequências de nucleotídeos (**Tabela 6**) e aminoácidos (**Tabela 7**) de quinze diferentes rhabdovírus dos gêneros *Nucleorhabdovirus* e *Cytorhabdovirus* completo depositados no bando de dados de DNA e proteína (GenBank/EMBL/DDBJ). As sequências das espécies SYNV e PYDV utilizados nos alinhamentos *pairwise* globais foram as de número de acesso NC_001615 e NC_016136, respectivamente.

Tabela 5. Identificação das sequências de aminoácidos das espécies virais *Sonchus yellow net nucleorhabdovirus* (SYNV) e *Potato yellow dwarf nucleorhabdovirus* (PYDV) utilizados nos alinhamentos *pairwise* globais.

Espécie – Acrônimo	Nº de Acesso	Proteína
<i>Sonchus yellow net nucleorhabdovirus</i> – SYNV	AAA50326.1	Proteína de movimento (sc4)
<i>Potato yellow dwarf nucleorhabdovirus</i> – PYDV	YP_004927966.1	X
<i>Potato yellow dwarf nucleorhabdovirus</i> – PYDV	YP_004927968.1	Proteína de movimento (Y)

Tabela 6. Sequências de nucleotídeos do genoma completo de vírus da família *Rhabdoviridae* depositadas no banco de dados (<https://www.ncbi.nlm.nih.gov/>) e utilizadas nas análises filogenéticas.

Espécie/Acrônimo	Nº de Acesso
<i>Alfalfa dwarf cytorhabdovirus</i> – ADV	NC_028237
<i>Barley yellow striate mosaic cytorhabdovirus</i> – BYSMV	NC_028244
<i>Datura yellow vein nucleorhabdovirus</i> – DYVV	NC_028231
<i>Eggplant mottled dwarf nucleorhabdovirus</i> – EMDV	NC_025389
<i>Lettuce necrotic yellows cytorhabdovirus</i> – LNYV	NC_007642
<i>Lettuce yellow mottle cytorhabdovirus</i> – LYMV	NC_011532
<i>Maize fine streak nucleorhabdovirus</i> – MFSV	NC_005974
<i>Maize iranian mosaic nucleorhabdovirus</i> – MIMV	NC_011542
<i>Maize mosaic nucleorhabdovirus</i> – MMV	NC_005975
<i>Northern cereal mosaic cytorhabdovirus</i> – NCMV	NC_002251
Persimmon virus A – PeVA	NC_018381
<i>Potato yellow dwarf nucleorhabdovirus</i> – PYDV	NC_016136
<i>Rice yellow stunt nucleorhabdovirus</i> – RYSV	NC_003746
<i>Sonchus yellow net nucleorhabdovirus</i> – SYNV	NC_001615
<i>Taro vein chlorosis nucleorhabdovirus</i> – TaVCV	NC_006942

Tabela 7. Sequências de aminoácidos das proteínas G (glicoproteína), N (nucleocapsídeo) e L (polimerase) de espécies virais da família *Rhabdoviridae* depositadas no banco de dados (<https://www.ncbi.nlm.nih.gov/>) e utilizados nas análises filogenéticas.

Espécie/Acrônimo	Nº de Acesso	Proteína
<i>Sonchus yellow net nucleorhabdovirus</i> - SYNV	NP_042285.1	
<i>Northern cereal mosaic cytorhabdovirus</i> - NCMV	NP_057961.1	
<i>Rice yellow stunt nucleorhabdovirus</i> - RYSV	NP_620500.1	
<i>Maize fine streak nucleorhabdovirus</i> - MFSV	YP_052848.1	
<i>Maize mosaic nucleorhabdovirus</i> - MMV	YP_052854.1	
<i>Taro vein chlorosis nucleorhabdovirus</i> - TaVCV	YP_224082.1	
<i>Lettuce yellow mottle cytorhabdovirus</i> - RYMV	YP_002308375.1	
<i>Maize Iranian mosaic nucleorhabdovirus</i> - MIMV	YP_002308458.1	Glicoproteína
<i>Potato yellow dwarf nucleorhabdovirus</i> - PYDV	YP_004927970.1	
<i>Eggplant mottled dwarf nucleorhabdovirus</i> - EMDV	YP_009094357.1	
<i>Datura yellow vein nucleorhabdovirus</i> - DYVV	YP_009176976.1	
<i>Alfalfa dwarf cyrorhabdovirus</i> - ADV	YP_009177019.1	
<i>Barley yellow striate mosaic cytorhabdovirus</i> - BYSMV	YP_009177229.1	

<i>Sonchus yellow net nucleorhabdovirus</i> - SYN	NP_042281.2	
<i>Northern cereal mosaic cytorhabdovirus</i> - NCM	NP_057954.1	
<i>Rice yellow stunt nucleorhabdovirus</i> - RYS	NP_620496.1	
<i>Maize fine streak nucleorhabdovirus</i> - MFS	YP_052843.1	
<i>Maize mosaic nucleorhabdovirus</i> - MM	YP_052850.1	
<i>Taro vein chlorosis nucleorhabdovirus</i> - TaV	YP_224078.1	
<i>Lettuce yellow mottle cytorhabdovirus</i> - RYM	YP_002308371.1	
<i>Maize Iranian mosaic nucleorhabdovirus</i> - MIM	YP_002308454.1	Nucleocapsídeo
<i>Potato yellow dwarf nucleorhabdovirus</i> - PYD	YP_004927965.1	
<i>Eggplant mottled dwarf nucleorhabdovirus</i> - EMD	YP_009094352.1	
<i>Datura yellow vein nucleorhabdovirus</i> - DYV	YP_009176972.1	
<i>Alfalfa dwarf cyrorhabdovirus</i> - AD	YP_009177015.1	
<i>Barley yellow striate mosaic cytorhabdovirus</i> - BYS	YP_009177222.1	
<i>Sonchus yellow net nucleorhabdovirus</i> - SYN	NP_042286	
<i>Northern cereal mosaic cytorhabdovirus</i> - NCM	NP_597914	
<i>Rice yellow stunt nucleorhabdovirus</i> - RYS	NP_620502	
<i>Maize fine streak nucleorhabdovirus</i> - MFS	YP_052849	
<i>Maize mosaic nucleorhabdovirus</i> - MM	YP_052855	
<i>Taro vein chlorosis nucleorhabdovirus</i> - TaV	YP_224083	
<i>Lettuce yellow mottle cytorhabdovirus</i> - RYM	YP_002308376	
<i>Lettuce necrotic yellows virus</i> - LNY	YP_425092	
<i>Maize Iranian mosaic nucleorhabdovirus</i> - MIM	YP_002308459	Polimerase
<i>Potato yellow dwarf nucleorhabdovirus</i> - PYD	YP_004927971	
<i>Eggplant mottled dwarf nucleorhabdovirus</i> - EMD	YP_009094358	
<i>Persimmon virus A</i> - PeV	YP_006576506	
<i>Datura yellow vein nucleorhabdovirus</i> - DYV	YP_009176977	
<i>Alfalfa dwarf cyrorhabdovirus</i> - AD	YP_009177021	
<i>Barley yellow striate mosaic cytorhabdovirus</i> - BYS	YP_009177231	

3. RESULTADOS E DISCUSSÃO

3.1. Plantas ornamentais coletadas no Viveiro I da NOVACAP (Companhia Urbanizadora da Nova Capital do Brasil)

As dez plantas ornamentais selecionadas foram coletadas no Viveiro I da NOVACAP e no momento da coleta apresentavam sintomas similares àqueles exibidos por plantas infectadas por vírus conforme ilustrado na **Figura 1**. Os sintomas observados foram mosaico, mosqueado, anéis cloróticos, amarelecimento e ápice encarquilhado (**Tabela 1**). A única amostra coletada que não fazia parte do sistema de produção da NOVACAP era o jasmim (*J. nitidum*), que fazia parte da composição paisagística do local.

3.2. Sequências obtidas via *Next Generation Sequencing* - NGS

O NGS do RNA total das amostras de plantas ornamentais resultou, primeiramente, em um total de 54.209.722 de pequenas sequências (*reads*), com tamanho de 101 pb. Estas, então, foram avaliadas pelo software *FastQC*, verificou-se que as sequências apresentaram boa qualidade por base, de acordo com o *software*, excetuando-se nas regiões terminais. Além disso, verificou-se a presença dos adaptadores. Para eliminar as extremidades dos *reads*, fez-se a trimagem utilizando o *software Trimmomatic*. Com esse programa, é possível, também, eliminar as sequências âncora, ou “adaptadores”. Utilizando, apenas, as condições e parâmetros padrões do programa, foi possível gerar novos *reads* sem adaptadores ou regiões de baixa cobertura, como analisado novamente pelo *FastQC*. Com a eliminação destas regiões sem cobertura e de adaptadores, o novo arquivo bruto “com cortes” passou a apresentar 45.449.068 *reads*. A adaptação das sequências para ajuste de qualidade é um fator importante para a boa continuidade nos demais ensaios de bioinformática (Bolger *et al.*, 2014).

As pequenas sequências foram montadas, resultando em porções maiores, os *contigs*. Com os *contigs*, passa a ser possível identificar a possível origem da sequência, via alinhamento local pelo algoritmo *tblastx*. Para isso, utilizou-se três diferentes montadores *Velvet*, *SPAdes* e *ABYSS* para se obter o máximo possível de *contigs* de qualidade, ou seja, em maior número e longos em tamanho. Como a programação dos montadores é continuamente alterada e adaptada às novas realidades da bioinformática e genômica, resultados podem variar dependendo do montador utilizado (**Tabela 8**). Há, portanto, uma variação entre *softwares* e isso pode comprometer a acurácia na detecção de patógenos em dado ambiente, como em uma planta, por exemplo. O mesmo montador pode variar conforme os parâmetros configurados, como com o “*k-mer*”. Este parâmetro refere-se a

todas as possíveis sequências, de tamanho k , de um *read*. Uma maneira de se avaliar e comparar os montadores é sua relação com *hits* com o *RefSeq* viral a partir do alinhamento local feito pelo *tblastx*. O resultado observado neste trabalho foi um baixo número de *hits* em geral, porém isso ocorre pelo fato de se ter utilizado um *RefSeq* puramente viral (e, ainda, sem bacteriófago) (**Tabelas 9**). No processo de sequenciamento via NGS, mesmo após o processo de semi-purificação e, portanto, enriquecimento de partículas virais, ácidos nucleicos de planta, bactéria e fungo podem ser, concomitantemente, sequenciados. Desta forma, muitos dos *contigs* montados, invariavelmente, não são de vírus e, sim, de outros organismos.

Tabela 8. Número de *contigs* gerados pelos montadores *SPAdes*, *Velvet* (*k-mer* 29 e *k-mer* 61) e *ABYSS*, número total de *hits* pelo algoritmo *tblastx* com base no *RefSeq* viral de setembro de 2016 para cada montador e número de *hits* específicos com vírus de planta e demais vírus, com base no resultado do algoritmo *tblastx* usando o *RefSeq* viral obtido em setembro de 2016.

	<i>SPAdes</i>	<i>Velvet k-mer 61</i>	<i>Velvet k-mer 29</i>	<i>ABYSS</i>
Contigs	95.736	78.406	1.657.028	2.882.857
Tamanho de contig (pb) – mínimo	72	121	57	19
Tamanho de contig (pb) – máximo	13.741	4.137	3.556	1.520
Hits	3.539	3.927	3.865	1.953
Vírus de planta	874	1.865	521	390
Outros vírus	2.665	2.062	3.344	1.563

Tabela 9. Quantidade de *hits* por família de vírus de planta, tamanho de *contigs*, identidade, cobertura e *E-value* a partir do algoritmo *tblastx* com base no RefSeq viral de setembro de 2016. Os *contigs* que geraram os *hits* foram produzidos pelos montadores *SPAdes*, *Velvet* (*k-mer* 61 e 29) e *ABYSS*.

Montador	Família de vírus	Quantidade de <i>hits</i>	Tamanho de <i>contigs</i> (pb) - mínimo	Tamanho de <i>contigs</i> (pb) - máximo	Identidade mínima (%)	Identidade máxima (%)	Cobertura mínima (%)	Cobertura máxima (%)	<i>E-value</i> menor	<i>E-value</i> maior
<i>SPAdes</i>	<i>Caulimoviridae</i>	213	22	491	27,4	100	7,85	100	0	8,91 e ⁻¹¹
	<i>Geminiviridae</i>	1	84	-*	50,0	-*	96,55	-*	1,27 e ⁻²⁵	-*
	<i>Luteoviridae</i>	6	40	467	60,8	77,5	1,09	25,54	0	3,37 e ⁻¹³
	<i>Potyviridae</i>	4	96	2.064	75,5	96,5	75,0	96,4	0	2,31 e ⁻⁴⁵
	<i>Rhabdoviridae</i>	9	107	1.708	40,3	95,8	21,95	99,52	0	4,17 e ⁻⁵⁸
	<i>Tombusviridae</i>	641	23	195	36,8	100	34,18	100	2,86 e ⁻⁸⁴	7,87 e ⁻¹¹
<i>Velvet (k-mer 61)</i>	<i>Betaflexiviridae</i>	1	56	-*	50,0	-*	89,36	-*	2,93 e ⁻¹²	-*
	<i>Caulimoviridae</i>	70	28	292	36,1	100	23,41	100	0	7,21 e ⁻¹¹
	<i>Closteroviridae</i>	1	46	-*	67,4	-*	69,70	-*	3,52 e ⁻¹⁶	-*
	<i>Luteoviridae</i>	12	38	481	41,7	84,2	23,12	99,17	0	7,18 e ⁻¹⁴
	<i>Potyviridae</i>	130	28	566	54,3	100	69,42	100	0	2,63 e ⁻¹²
	<i>Rhabdoviridae</i>	124	22	962	34,0	100	19,41	100	0	9,11 e ⁻¹¹
<i>Velvet (k-mer 29)</i>	<i>Tombusviridae</i>	1527	35	203	41,1	95,0	35,15	100	5,22 e ⁻¹²³	9,84 e ⁻¹¹
	<i>Betaflexiviridae</i>	1	58	-*	50,0	-*	97,15	-*	9,62 e ⁻¹³	-*
	<i>Caulimoviridae</i>	358	22	187	26,8	100	22,6	100	7,97 e ⁻¹²⁸	9,89 e ⁻¹¹
	<i>Closteroviridae</i>	1	85	-*	55,3	-*	78,46	-*	9,38 e ⁻²⁸	-*
	<i>Geminiviridae</i>	1	99	-*	44,4	-*	98,34	-*	1,22 e ⁻²⁴	-*
	<i>Luteoviridae</i>	12	40	111	41,4	78,7	39,63	99,46	7,19 e ⁻⁴⁹	8,02 e ⁻¹¹
<i>ABYSS</i>	<i>Partitiviridae</i>	11	32	89	46,7	80,5	68,46	100	1,51 e ⁻³⁷	6,27 e ⁻¹²
	<i>Potyviridae</i>	30	25	72	64,2	100	92,73	99,13	1,64 e ⁻⁴⁴	7,32 e ⁻¹¹
	<i>Rhabdoviridae</i>	98	23	172	34,4	100	33,64	100	1,06 e ⁻⁷⁶	9,42 e ⁻¹¹
	<i>Tombusviridae</i>	6	31	205	63,8	77,4	49,32	98,28	3,07 e ⁻¹⁰⁶	3,05 e ⁻¹²
	<i>Tymoviridae</i>	3	40	51	78,4	90,0	96,0	99,35	5,66 e ⁻²¹	1,41 e ⁻¹⁹
	<i>Caulimoviridae</i>	144	23	116	46,7	100	46,28	100	4,45 e ⁻⁹⁰	7,32 e ⁻¹¹
<i>ABYSS</i>	<i>Closteroviridae</i>	1	33	-*	75,8	-*	93,40	-*	4,54 e ⁻¹³	-*
	<i>Geminiviridae</i>	1	68	-*	44,1	-*	76,98	-*	2,30 e ⁻¹⁵	-*
	<i>Luteoviridae</i>	17	36	277	39,3	83,3	41,10	97,98	9,60 e ⁻¹²⁴	2,77 e ⁻¹¹
	<i>Potyviridae</i>	23	23	63	81,0	100	95,83	100	4,14 e ⁻³⁷	7,32 e ⁻¹¹
	<i>Rhabdoviridae</i>	185	25	154	35,4	100	50,42	100	2,21 e ⁻⁹⁵	5,59 e ⁻¹¹
	<i>Tombusviridae</i>	17	31	89	39,7	81,8	39,57	100	1,99 e ⁻⁴⁹	1,51 e ⁻¹¹

* Não se aplica onde só há um *contig*.

A avaliação da qualidade dos montadores quanto ao tipo de vírus, ou seja, se ele é um vírus de planta ou não, é feita eliminando artefatos de planta, bactérias e fungos do sequenciamento, já que o que é avaliado está dentro do *hit* mostrado na **Figura 3** (Bao *et al.*, 2011). Nota-se que os parâmetros são importantes, como no caso do *k-mer* no montador *Velvet*. A variedade entre montadores pode ser, também, devido a quantidade de *contigs* produzido por cada um, como visto pelos 2.882.857 *contigs* pelo *ABYSS* e os 78.406 *contigs* pelo *Velvet* (*k-mer* 61) (**Tabela 8**). A alta quantidade de sequências, mesmo que montadas, pode diluir os resultados esperados e, portanto, dificultar os trabalhos de bioinformática de análise e interpretação.

Os montadores apresentaram, também, uma diferença significativa na presença e quantidade de sequências de diferentes famílias de vírus de plantas (**Figura 4**). Isso pode gerar falsos negativos e positivos, já que alguns casos, como para a família *Geminiviridae* apenas um *contig* foi observado. A identificação de vírus de DNA para um NGS feito com base em RNA total pode ser devido a presença de RNAs mensageiros (RNAm) destes vírus de DNA.

Constatou-se uma relação em número para alguns grupos de vírus no NGS, independente do montador utilizado. *Hits* com vírus das famílias como *Tombusviridae*, *Potyviridae*, *Rhabdoviridae* e *Caulimoviridae* foram predominantes na lista de *contigs* com identidade viral. Isso pode oferecer evidências em relação à presença desses vírus nas amostras. Isso foi comprovado para a família *Rhabdoviridae*, onde pelo menos duas prováveis espécies consideradas novas foram detectadas nas amostras *C. lanceolata* e *P. lutea*. Neste trabalho, por questão de tempo, não foram montados genomas de sequências com *hit* para a família *Caulimoviridae* e demais famílias.

O uso de NGS associado a ferramentas de bioinformática também é utilizado na detecção de vírus causando doenças em espécies cultivadas. Adams *et al.* (2013), por exemplo, identificaram uma espécie da família *Tombusviridae* infectando milho em diferentes regiões de Quênia. A partir da bioinformática, foi possível montar uma sequência com genoma completo do machlomovírus *Maize chlorotic mottle virus* – MCMV (*Tombusviridae*). O sequenciamento retornou com *reads* de vírus de planta equivalendo a 4% do total (Adams *et al.*, 2013). Neste trabalho, a partir do montador *SPAdes*, o NGS retornou com *hits* de vírus equivalendo a 3,69% do total.

Para uma espécie de crisântemo (*Chrysanthemum zawadskii* var. *latilobum*), Seo *et al.* (2015) realizaram uma análise com uso de NGS visando identificar possíveis agentes causais de sintomas típicos de vírus, como mosaico e amarelecimento observados nas

plantas. As análises de bioinformática resultaram em diferentes *contigs*. Um *contig*, de 4.617 pb, retornou do algoritmo *tblastx* como uma possível sequência do isolado Gyp1 do vírus *Carnation Italian ringspot virus* (CIRV). O CIRV pertence ao gênero *Tombusvirus* (*Tombusviridae*) e foi primeiramente descrito na planta ornamental *Dianthus caryophyllus* (craveiro). O isolado Gyp1 foi descrito em *Gypsophila paniculata* (mosquitinho-branco), outra planta ornamental. Ainda de acordo com Seo *et al.* (2015), o CIRV identificado em *C. zawadskii* pode servir como fonte de inóculo para plantas ornamentais produzidas na região, como craveiro, mosquitinho-branco e outras espécies de crisântemo (Seo *et al.*, 2015).

Novas informações por meio de NGS também vem sendo adquiridas para o gênero *Potyvirus*, principalmente na identificação e classificação de novos isolados. Novas sequências de *Bean yellow mosaic virus* (BYMV) foram identificadas e montadas a partir de amostras de *Lupinus angustifolius* (Fabaceae). A partir do sequenciamento, foram identificadas vinte e três possíveis novas sequências de isolados de BYMV. A partir do mapeamento, estas sequências tiveram o seu genoma completo montado e comparado aos demais isolados da espécie viral. Com as análises filogenéticas entre as sequências de genomas montados a partir do NGS com as demais depositadas em bancos de dados, foi possível propor novos critérios taxonômicos para o complexo BYMV (Kehoe *et al.*, 2014).

Zhai *et al.* (2014), por meio de NGS, detectaram o nucleorhabdovírus *Eggplant mottled dwarf nucleorhabdovirus* (EMDV) na planta ornamental agapanto (*Agapanthus africanus*), uma planta ornamental. A sequência do vírus foi completamente montada e teve seu genoma completo, incluindo os terminais 3' e 5', sequenciado via método Sanger (Zhai *et al.*, 2014). O vírus EMDV também já foi relatado em hibisco (*Hibiscus rosa-sinensis*) (De Stradis *et al.*, 2008).

Neste trabalho, diferentes *softwares*, inclusive muitos disponíveis de maneira aberta (*open access*), resultaram em diferentes conjuntos de dados que, juntos, podem assegurar um dado final de maior qualidade. A escolha dos *softwares*, também, é tão importante quanto a definição dos parâmetros usados. Diferentes montadores, por exemplo, geram diferentes resultados e, portanto, novos dados que devem ser analisados e interpretados. O montador *Velvet* apresentou alta diferença entre os *hits* para a família *Tombusviridae* entre os parâmetros do *k-mer*, com 1527 *hits* para o *k-mer* 61 e 6 *hits* para o *k-mer* 29. Além da diferença, o montador *Velvet* com *k-mer* 61 apresentou um número menor de *contigs*.

Estudos já avaliaram e compararam diferentes ferramentas utilizadas na montagem *de novo* de sequências para produtos oriundos de NGS. Bao *et al.* (2011) avaliaram, entre outros montadores, os *softwares* *ABYSS* (versão 1.2.6) e *Velvet* (versão 1.0.09). Ambos

apresentaram um uso da memória RAM em um valor próximo a 9 *Gigabytes* (GB). Também apresentaram o tamanho mínimo de *contig* aproximado: 25 pb para o *ABYSS* e 21 pb para o *Velvet*. Para os outros critérios, os dois montadores apresentaram resultados discrepantes. Para o tamanho máximo de *contig*, o *ABYSS* resultou em 12.804 pb, enquanto que o *Velvet* 2.285 pb. A cobertura do genoma foi maior para o *Velvet*, com 17,47%. O montador *ABYSS* teve uma cobertura de 5,95%. Apesar disso, o *Velvet* apresentou um tempo maior de processamento, com 357,26 minutos; o *ABYSS*, por sua vez, 31 minutos. Os resultados discrepantes entre *ABYSS* e *Velvet* refletem para os demais montadores utilizados (Bao et al., 2011). Estes resultados corroboram com a ideia de que diferentes resultados podem ser produzidos por diferentes programas.

Apesar da combinação de programas se mostrar uma alternativa para a busca por diferentes resultados, programas com fluxogramas completos e específicos estão disponíveis para um uso mais rápido e cotidiano do NGS. Alguns desses programas, como o *VirFind*, já apresentam parâmetros definidos para trabalhar em sequenciamentos virais. Esse tipo de metodologia pode auxiliar na diagnose de doenças (Ho & Tzanetakis, 2014).

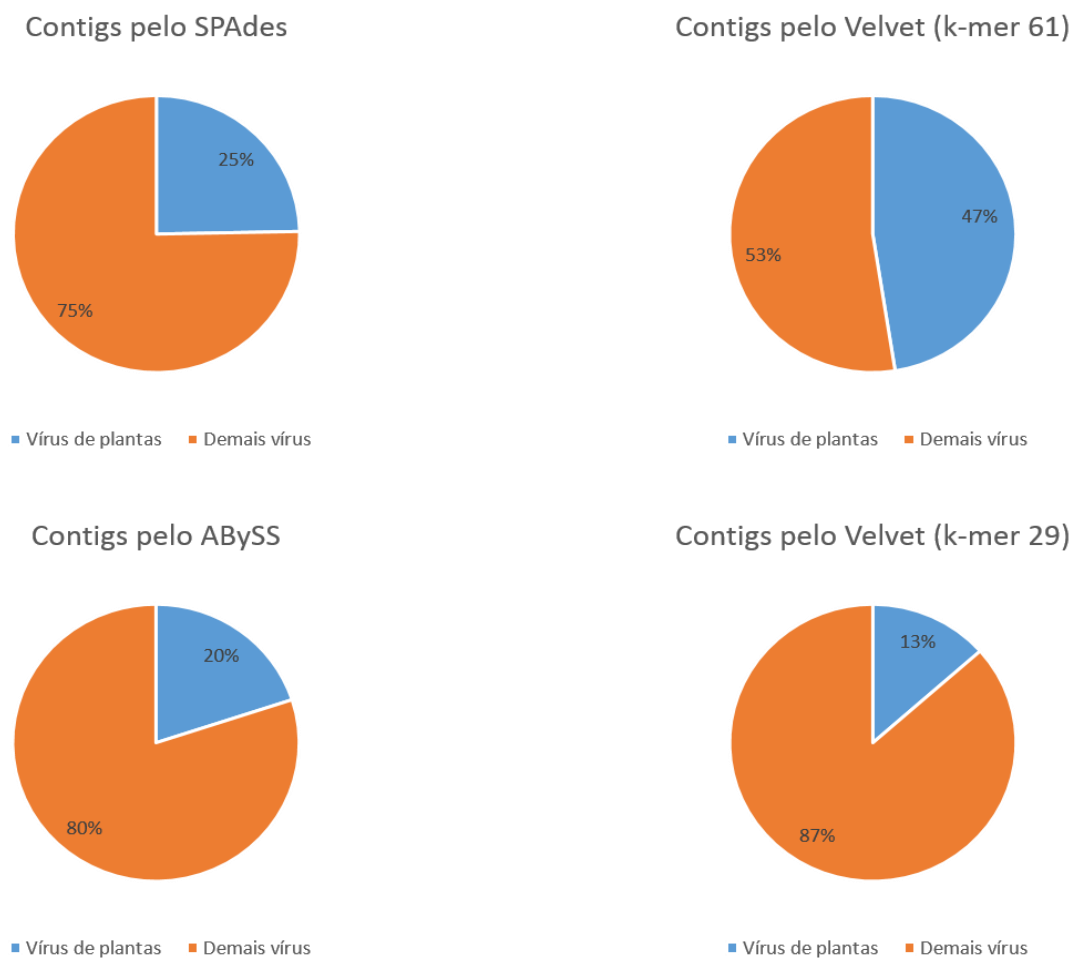


Figura 3. Gráficos comparativos entre os montadores de sequências (*SPAdes*, *Velvet* e *ABySS*) ilustrando a relação entre vírus de planta e os demais vírus obtidos por meio de sequenciamento NGS (*Next-Generation Sequencing*), utilizando sequências que apresentaram *hits* com sequências virais.

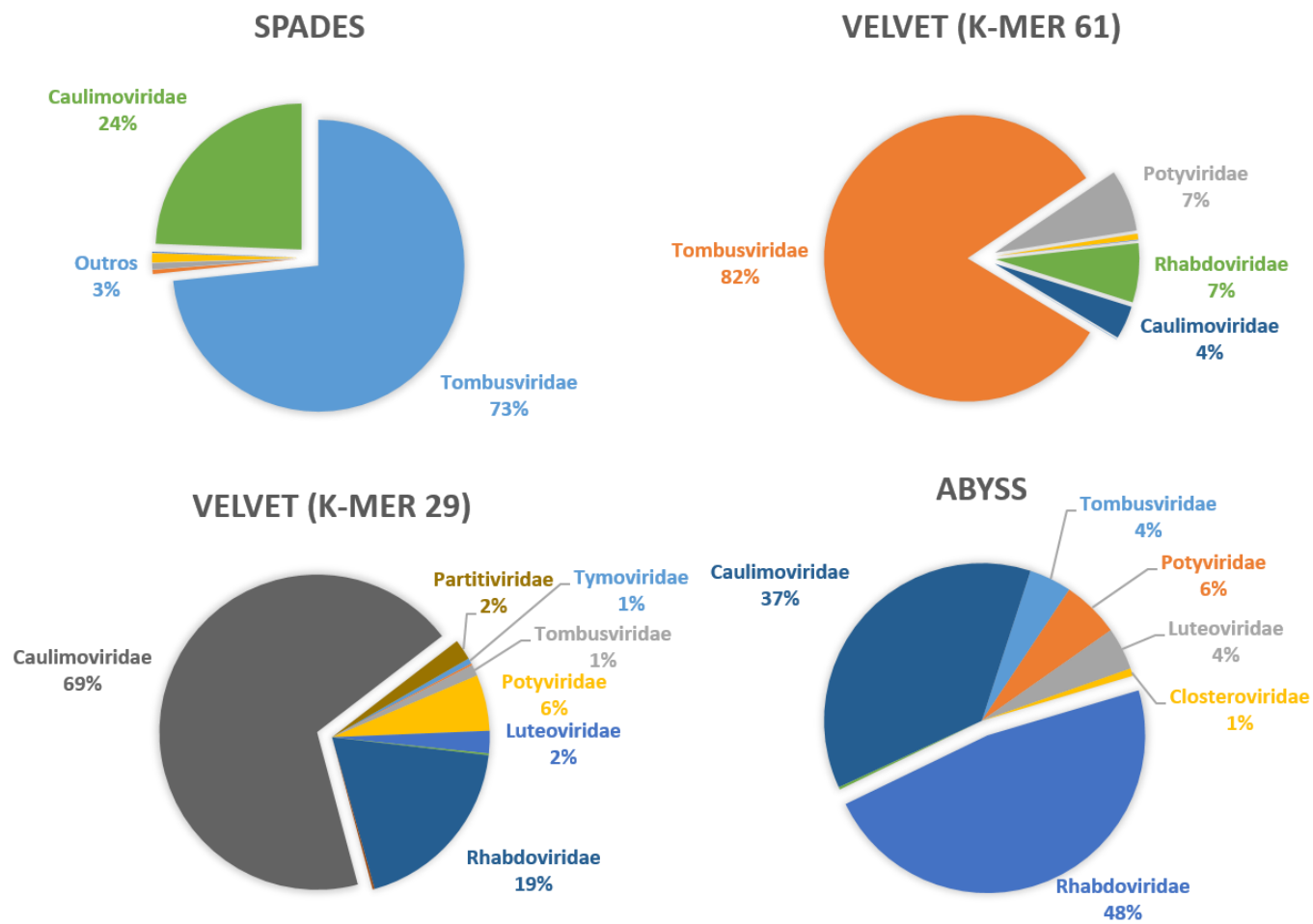


Figura 4. Gráficos comparativos entre os montadores de sequências *SPAdes*, *Velvet* e *ABySS* referentes a presença de vírus pertencentes a determinadas famílias de vírus de plantas e sua relação com o número total de vírus de planta. Os representantes que não ultrapassaram a margem de 0% não estão apontados nos gráficos.

3.3. Genomas montados e seleção das espécies *Coreopsis mottle-associated virus* (CMAV) e *Pachystachys mosaic-associated virus* (PMAV)

Com os *contigs* devidamente montados e divididos, fez-se o alinhamento local pelo algoritmo *tblastx* usando como referência o *RefSeq* viral sem bacteriófago obtido em setembro de 2016 (5.495 sequências). Com essa metodologia, separa-se genomas de plantas, bactérias, fungos, além de mRNA de plantas, entre outros. Especificar a natureza viral, no entanto, não significa que os resultados serão claros. Muitos dos vírus detectados infectam apenas célula animal, o que pode representar um artefato ou contaminação, devido aos *contigs* de pequeno tamanho ou de baixa identidade. Foram trabalhadas sequências com identidade maior que 50% com os genomas de referência obtidos via *tblastx*, já que a probabilidade destas serem um artefato é menor. Desta forma, elimina-se as sequências consideradas “virais”, inclusive no grupo de vírus de plantas. Sequências de genoma viral que apareceram nos resultados podem ser artefatos, mediante sua baixa (2-10 *contigs* por montador) ou baixíssima (1) presença (**Figura 4**).

Para dar continuidade aos trabalhos de bioinformática e, conseqüentemente, moleculares, selecionou-se seis diferentes grupos de *contigs*. Esses grupos apresentaram 500 a 1.000 *contigs* e/ou *contigs* longos (maiores que 3.000 pb). *Contigs* longos são especialmente importantes para a elaboração e montagem do genoma completo do possível vírus ali representado. O montador *SPAdes*, por exemplo, foi capaz de montar *contigs* que ultrapassaram 1.000 pb, o que facilitou, em muito, completar a viral computacionalmente. O *ABYSS*, em contrapartida, produziu um número maior de *contigs* em relação aos demais montadores, mas com baixo índice de *hits* com sequências virais (**Tabela 8**).

Com base nos parâmetros já mencionados, os grupos de *hits* selecionados foram, com base no genoma de referência de três membros da família *Rhabdoviridae* [*Sonchus yellow net nucleorhabdovirus* (SYNV), *Potato yellow dwarf nucleorhabdovirus* (PYDV) e *Alfalfa dwarf cytorhabdovirus* (ADV)]; duas espécies da família *Tombusviridae*, *Rosa rugosa leaf distortion virus* (RrLDV) e *Carrot mottle mimic virus* (CMoMV); e uma da família *Potyviriidae* *Jasmine ringspot virus* (JRSV)]. A escolha destes grupos virais (e as famílias do genoma de referência) condiz com o dado mostrado pela **Figura 4**. A alta quantidade de *hits* de *Tombusviridae* resultou em uma montagem de duas sequências compatíveis com genoma viral, enquanto que em *Rhabdoviridae* foram três. Para a família *Potyviriidae*, também bem representado nos gráficos da **Figura 4**, um genoma foi recuperado. A relação das sequências montadas com os respectivos *contigs* é representada pela Tabela 10.

Tabela 10. Quantidade de *contigs*, tamanho mínimo e máximo (em pb) e *E-value* menor e maior para cada sequência viral trabalhada *in silico* – *Sonchus yellow net nucleorhabdovirus* (SYNV), *Potato yellow dwarf nucleorhabdovirus* (PYDV), *Alfalfa dwarf cytorhabdovirus* (ADV), *Jasmine ringspot virus* (JRSV), *Carrot mottle mimic virus* (CMoMV) e *Rosa rugosa leaf distortion virus* (RrLDV) em cada montador utilizado (*SPAdes*, *Velvet* e *ABYSS*).

Montador	Sequência viral	Quantidade de <i>contigs</i>	Tamanho de contig (pb) - mínimo	Tamanho de contig (pb) - máximo	<i>E-value</i> menor	<i>E-value</i> maior
<i>SPAdes</i>	SYNV	5	107	1.708	0	4,17 e ⁻⁵⁸
	PYDV	1	962	-	0	-
	ADV	3	238	1.122	0	2,50 e ⁻⁸⁵
	JRSV	2	96	1.321	0	2,31 e ⁻⁴⁵
	CMoMV	0	-	-	-	-
	RrLDV	339	24	195	5,31 e ⁻¹¹	3,42 e ⁻⁸³
<i>Velvet (k-mer 61)</i>	SYNV	85	22	251	0	1,79 e ⁻¹³
	PYDV	13	40	962	0	1,46 e ⁻¹⁴
	ADV	26	29	421	0	9,11 e ⁻¹¹
	JRSV	95	28	244	1,02 e ⁻¹⁴⁶	2,63 e ⁻¹²
	CMoMV	3	66	108	5,13 e ⁻²⁹	2,77 e ⁻²⁹
	RrLDV	535	35	60	1,92 e ⁻²⁷	8,60 e ⁻¹¹
<i>Velvet (k-mer 29)</i>	SYNV	43	23	96	2,14 e ⁻⁵⁷	9,42 e ⁻¹¹
	PYDV	23	27	118	4,45 e ⁻⁵⁰	2,73 e ⁻¹¹
	ADV	30	30	172	1,06 e ⁻⁷⁶	7,32 e ⁻¹¹
	JRSV	13	28	71	1,24 e ⁻³⁴	5,76 e ⁻¹²
	CMoMV	5	43	205	3,07 e ⁻¹⁰⁶	9,26 e ⁻¹⁸
	RrLDV	0	-	-	-	-
<i>ABYSS</i>	SYNV	79	25	154	2,21 e ⁻⁹⁵	4,39 e ⁻¹¹
	PYDV	58	58	66	2,24 e ⁻³⁶	3,36 e ⁻¹¹
	ADV	48	50	151	5,05 e ⁻⁷⁰	5,59 e ⁻¹¹
	JRSV	17	23	63	4,14 e ⁻³⁷	7,32 e ⁻¹¹
	CMoMV	5	40	79	1,99 e ⁻⁴⁹	6,23 e ⁻¹⁷
	RrLDV	0	-	-	-	-

A montagem do genoma da sequência de referência PYDV resultou em um mapeamento com 115.038 *reads* com uma cobertura máxima de 2.366 bases e média de 1.009 bases. O uso da sequência de genoma de referência SYNV resultou em um mapeamento total de 85.008 *reads*, com cobertura máxima de 1.508 bases e média de 614 bases. O mapeamento da sequência obtida a partir do genoma de referência JRSV retornou em um total de 101.957 *reads*, com cobertura máxima de 2.518 bases e média de 1060 bases. As três sequências, PYDV, SYNV e JRSV, tiveram a sequência do genoma completo a partir do mapeamento dos *reads*, usando o genoma de referência como parâmetro. As sequências compatíveis com aquelas referências ADV, CMoMV e RrLDV foram montadas parcialmente. A sequência ADV foi mapeada com 899 *reads*, com uma cobertura máxima

de 41 bases e média de 18 bases. Para a sequência CMoMV, a montagem foi feita com 8.237 *reads*, com máxima de 2.525 bases e média de 225 bases. A sequência RrLDV foi mapeada com 9.258.824 *reads*, com cobertura máxima de 1.209.705 bases e média de 291.297 bases.

Os resultados deste trabalho são semelhantes aos de Zhai *et al.* (2014) realizado na Coreia. Neste, foi feita a montagem da sequência do genoma completo do nucleorhabdovírus EMDV a partir de 14 milhões de *reads* oriundos de um NGS feito a partir de agapanto. A montagem *de novo* resultou em 33.000 *contigs*. O maior, apresentando 13,1 kb, foi mapeado por 5,9 milhões de *reads* (Zhai *et al.*, 2014).

3.4. Detecção das sequências virais obtidas via *Next Generation Sequencing* – NGS no *pool* e nas amostras de plantas ornamentais individuais

A região escolhida para o ensaio de detecção foi a provável região da polimerase, identificada por meio de comparação com os respectivos genomas de referência, para todos os grupos. Identificar as espécies virais usando *primers* sintetizados com base nas sequências obtidas via NGS a partir de um *pool* ou de amostras individuais é um forte alicerce para que aquele genoma recuperado por meio de NGS esteja, de fato, representando um vírus presente na amostra.

Neste trabalho, os *primers* sintetizados (**Tabela 2**) para amplificação de DNA específico da região da polimerase foram utilizados para confecção de cDNAs a partir do RNA total tanto do *pool* de amostras (alíquota do mesmo material enviado ao NGS) quanto das amostras de plantas individuais. Inicialmente, *amplicons* de tamanho esperado foram observados quando se utilizou os seis pares de *primers* acima citados para detecção viral a partir do *pool*, o qual, após este resultado, passou a ser utilizado como controle positivo de todas as PCR realizadas nas amostras individuais. Exceto, quando se usou o par de *primers* para o tombusvírus RrLDV, nenhuma amplificação foi observada nos controles.

Os rhabdovírus SYNIV e PYDV foram detectados no *pool* de amostras, assim como as demais sequências, que também foram positivas nas reações de amplificação a partir do *pool*. Para as detecções em amostras individuais, resultados positivos foram observados para PYDV na amostra 3 (*P. lutea* – camarão-amarelo), SYNIV e CMoMV na amostra 9 (*C. lanceolata* – margaridinha) e RrLDV na amostra 10 (*J. nitidum* – jasmim-estrela). Para JRSV e ADV nenhum resultado positivo foi observado em nenhuma das amostras individuais testadas.

O fato do teste realizado com *primers* específicos para ADV e JRSV resultarem em amplificação positiva apenas quando utilizou-se como molde RNA proveniente do *pool* das

amostras indica que pode haver problemas com a qualidade do RNA total extraído das amostras individuais, embora o protocolo utilizado tenha sido o mesmo. Como a extração de RNA foi realizada diretamente do tecido da planta, há chances de co-purificação de contaminantes que podem inibir a reação de transcrição reversa ou PCR, além de possíveis diferenças na população viral entre as folhas usadas para extração de RNA total do *pool* e extração de RNA total individual. Como todas as PCR foram realizadas seguindo o mesmo programa, houve amplificação de fragmentos de DNA de tamanho não esperado, como observado quando se utilizou os *primers* específicos para SYNV e PYDV.

3.5. Determinação da sequência do genoma completo das espécies *Coreopsis mottle-associated virus* (CMAV) e *Pachystachys mosaic-associated virus* (PMAV)

As bandas representando os *amplicons* das sequências SYNV e PYDV foram selecionadas para os ensaios posteriores. Os vírus com identidade ao SYNV e ao PYDV foram nomeados tentativamente como *Coreopsis mottle-associated virus* (CMAV) e *Pachystachys mosaic-associated virus* (PMAV), respectivamente. O vírus CMAV foi detectado na amostra de *Coreopsis lanceolata* (margaridinha) e PMAV na *Pachystachys lutea* (camarão-amarelo). Como critério de nomeação utilizou-se a hospedeira na qual a espécie viral foi detectada acrescida da palavra *associated*. A escolha do “associated” no nome é devido a não realização dos postulados de Koch para provar efetivamente a relação de doença.

Os dois vírus foram selecionados para uma caracterização detalhada, incluindo a determinação da sequência do genoma completo via Sanger a partir da planta/amostra. Essa estratégia foi utilizada para comprovar com um método alternativo a presença desses vírus nas amostras. Ainda, no método de construção de genomas *in silico*, não há um “limite” na alongação do mesmo; ou seja, uma sequência pode estar maior ou menor que a “real”. Para tal, sequenciar as regiões terminais 3’ e 5’ se mostra crucial e o método RACE é o indicado, assim como feito para recuperar o genoma completo de um novo vírus do gênero *Marafivirus* (*Tymoviridae*) em parreira (*Vitis vinifera*) (Al Rwahnih et al., 2009).

Para a obtenção do genoma completo de CMAV, foram desenhados *primers* em cinco diferentes regiões do genoma (1, 2, 3, 4 e 5) e os fragmentos de DNA resultantes da amplificação foram de aproximadamente 2.900 pb conforme ilustrado na **Figura 5**. Posteriormente, o sequenciamento Sanger confirmou as sequências. Para o PMAV, os *primers* utilizados para a amplificação por PCR das 4 regiões do genoma (1, 2, 3 e 4) resultaram em *amplicons* esperados de cerca de 2.900 pb para as regiões 1, 2 e 4 e 3.000 pb

para a região 3, como ilustrado na **Figura 6**. Em seguida, o sequenciamento Sanger também confirmou as quatro sequências.

Similarmente, Higgins et al. (2016) detectaram um novo cytorhabdovírus, *Colocasia bobone disease-associated virus* (CBDAV), em inhame (*Colocasia esculenta*), e recuperaram a sequência completa do genoma PCR e RACE para os terminais 3' e 5'.

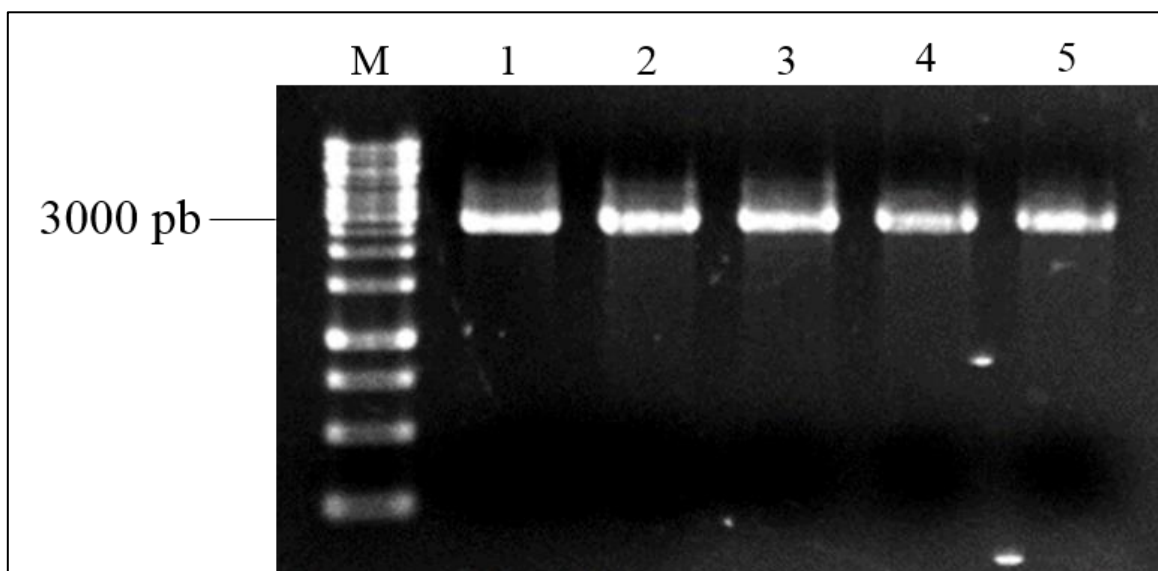


Figura 5. Eletroforese em gel de agarose (1%) da amplificação de diferentes regiões do genoma de *Coreopsis mottle-associated nucleorhabdovirus* (CMAV), denominadas 1, 2, 3, 4 e 5, a partir de amostras de *Coreopsis lanceolata* (margaridinha). Cinco pares de *primers* (um para cada região de tamanho aproximado de 2.900 pb) foram utilizados para cobrir o genoma de 13.657 pb. M: Marcador Molecular 1 kb Plus (Invitrogen) evidenciado na primeira coluna. Tamanho esperado do *amplicon*: 2.900 pb.

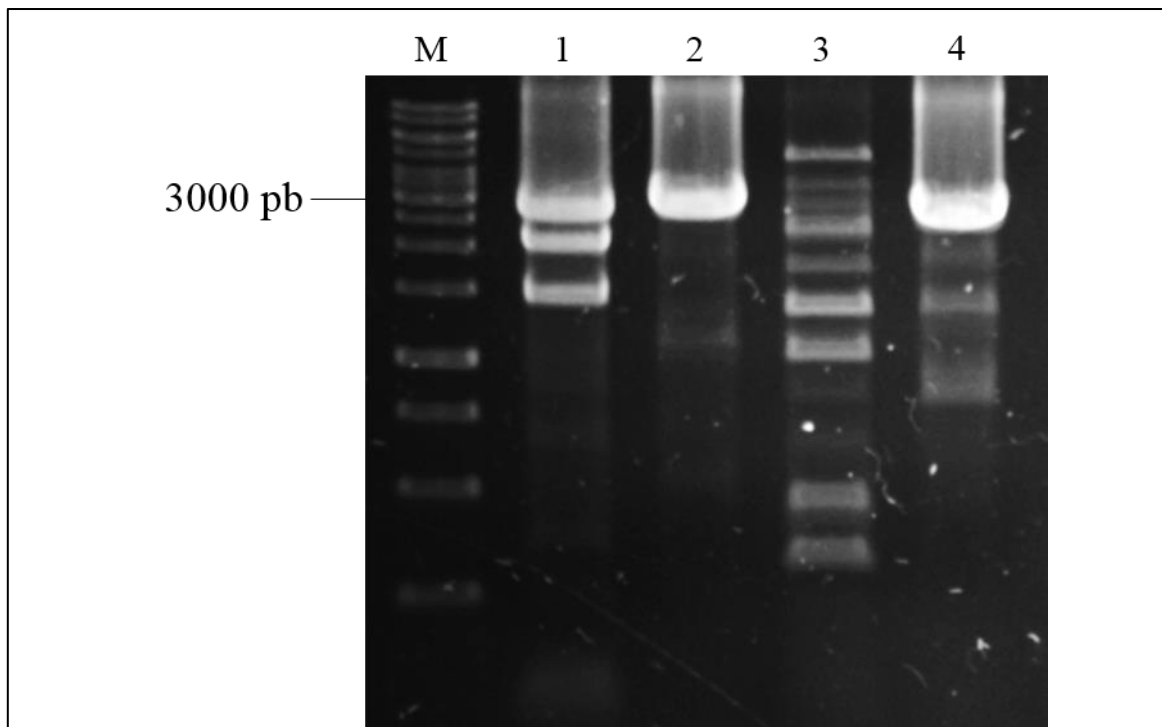


Figura 6. Eletroforese em gel de agarose (1%) da amplificação de diferentes regiões do genoma de *Pachystachys* mosaic-associated *nucleorhabdovirus* (PMaV), denominados 1, 2, 3 e 4, a partir de amostras de *Pachystachys lutea* (camarão-amarelo). Quatro pares de *primers* (um para cada região de tamanho aproximado de 2.900 pb) foram utilizados para cobrir o genoma de 11.342 pb. M: Marcador Molecular 1 kb Plus (Invitrogen) na primeira coluna. Tamanho esperado do *amplicon*: 2.900 pb.

Os fragmentos (1, 2, 3, 4 e 5 para CMaV e 1, 2, 3 e 4 para PMaV) representavam o genoma completo de ambos os vírus, excetuando-se as regiões terminais 3' e 5'. Para isso, foi feito o protocolo 3'RACE e 5'RACE para os dois vírus. Os produtos da amplificação das PCRs dos protocolos de 3' e 5' RACE foram eluídos, purificados e sequenciados via método Sanger.

O sequenciamento Sanger para o genoma completo de CMaV e PMaV, recuperados diretamente das plantas *Coreopsis lanceolata* e *Pachystachys lutea*, resultou em sequências com identidade nucleotídica maior que 95% com as obtidas e montadas via NGS e Bioinformática. Isto é, assumindo-se que o sequenciamento Sanger, assim como a de NGS, ocorre mediante um agrupamento de populações virais, acredita-se que a sequência final seja uma “média” em relação a polimorfismos e posição de SNPs (*Single nucleotide polymorphisms*). Diferenças foram observadas nas regiões terminais. Para o CMaV, houve uma redução 140 pb em relação ao NGS no total. Para o PMaV, foram 251 pb a menos.

O genoma de CMaV apresentou um tamanho de 13.657 pb e é significativamente próximo do SYNIV, de 13.720 pb, apresentando, no entanto, diferenças em todas as regiões intergênicas. Enquanto que no SYNIV essa região apresenta 2 pb, a do CMaV ultrapassa 300

pb cada. O restante, no entanto, segue os padrões do SYNV, como a posição e orientação dos genes e ORFs. Por analogia, pensa-se que o genoma do CMaV apresente os mesmos genes de SYNV (**Figura 7**). Típico de rhabdovírus, o CMaV também apresentou uma sequência repetitiva: GUUGGUUUUCUUA. Essa sequência foi encontrada entre todas as ORFs. A sequência repetitiva de SYNV é UAUAUUCUUUUU (número de acesso NC_001615 no GenBank - <https://www.ncbi.nlm.nih.gov/genbank/>).

Coreopsis mottle-associated virus (CMaV)

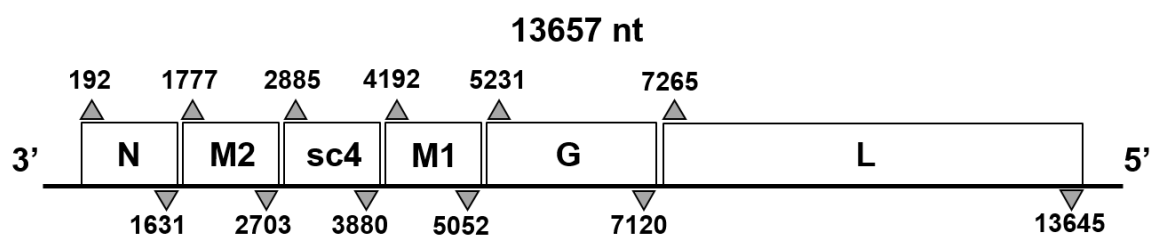


Figura 7. Esquema com base na sequência obtida via sequenciamento Sanger representando o genoma do Coreopsis mottle-associated virus (CMaV) e seus possíveis genes. As setas indicam a posição do início e fim de cada gene.

O genoma de PMaV, por sua vez, apresentou um genoma com 11.342 pb e 7 possíveis ORFs a partir do sequenciamento Sanger (**Figura 8**). O PMaV apresenta regiões intergênicas distintas entre cada ORF e, por isso, não foi possível identificar uma sequência repetitiva. Os genes putativos foram determinados com base na analogia ao PYDV.

Pachystachys mosaic-associated virus (PMaV)

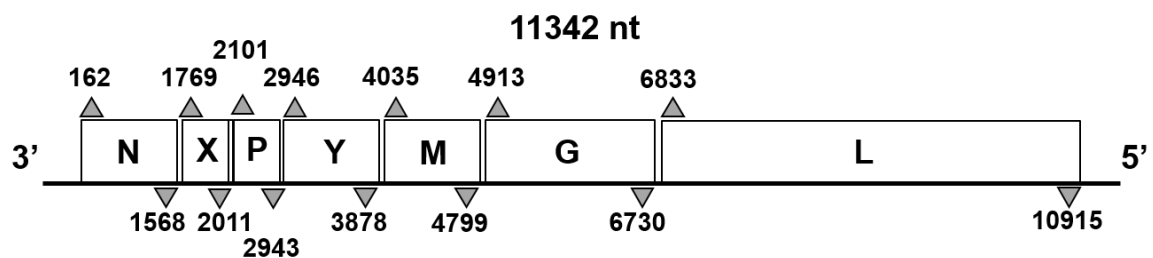


Figura 8. Esquema com base na sequência obtida via sequenciamento Sanger representando o genoma do Pachystachys mosaic-associated virus (PMaV) e seus possíveis genes. As setas indicam a posição do início e fim de cada gene.

A recuperação do genoma, além das suas regiões terminais 3' e 5', é importante para comprovar as sequências obtidas via NGS. Nesse trabalho, o genoma foi montado *in silico* e os terminais 3' e 5' recuperados via protocolo RACE seguido de sequenciamento Sanger.

Similarmente, Zhai *et al.* (2014) identificaram um novo isolado de EMDV, montado e recuperado a partir de NGS em agapanto. A partir da detecção e da recuperação do genoma, a planta ornamental passou a ser considerada uma nova hospedeira para o nucleorhabdovírus EMDV (Zhai *et al.*, 2014).

Com o sequenciamento do método Sanger, foi possível verificar que os genomas montados *in silico* apresentaram alta identidade (95%) nucleotídica com os vírus que se encontravam presentes nas amostras *C. lanceolata* e *P. lutea* e sequenciados via método Sanger. Para assegurar se dado vírus é ou não espécie nova de dado gênero, cabe a análises mais profundas do genoma, como pela análise de alinhamentos *pairwise*, múltiplos e árvores filogenéticas.

3.6. Filogenia das sequências virais Coreopsis mottle-associated virus (CMaV) e Pachystachys mosaic-associated virus (PMaV)

O alinhamento múltiplo de sequências global (MSA) e *pairwise* do genoma do CMaV com diferentes outros rhabdovírus, inclusive *Cytorhabdovirus*, mostrou a relação do vírus com os demais integrantes da família (**Figura 9**). O vírus CMaV ficou apresentou maior identidade ao SYNV, seguido do *Datura yellow vein nucleorhabdovirus* (DYVV). O alinhamento do genoma de CMaV com o genoma de SYNV resultou em uma identidade a nível de nucleotídeo de 56,99%. Para os demais vírus, inclusive do gênero *Nucleorhabdovirus*, a identidade é inferior a 40% em relação ao CMaV.

Em nível de proteína, o CMaV apresentou uma alta identidade com o SYNV nas proteínas L, N, G e sc4. A proteína L (polimerase) de CMaV apresentou 87,15% de identidade com a proteína L de SYNV. A segunda maior identidade foi com o DYVV, com 44,72% de identidade (**Figura 10**). A proteína N (nucleocapsídeo) de CMaV apresentou 93,26% de identidade com a proteína N de SYNV e 50,00% de identidade com a proteína N de DYVV (**Figura 11**). Para a proteína G (glicoproteína), a identidade foi de 94,59% e 41,88% para a proteína G de SYNV e DYVV, respectivamente (**Figura 12**). O alinhamento *pairwise* da proteína de movimento sc4 entre CMaV e SYNV resultou em uma identidade de 94%.

As árvores filogenéticas com base no genoma (a nível de nucleotídeo) (**Figura 13**) e proteínas N, G e L (aminoácido) mostram a divisão entre os gêneros *Nucleorhabdovirus* e *Cytorhabdovirus*. Além disso, evidencia-se a relação entre o CMaV e o SYNV. O DYVV é o segundo vírus mais próximo de CMaV. Os resultados, também, indicam que o vírus faz parte do gênero *Nucleorhabdovirus*, juntamente com os demais representados (**Figuras 14,**

15 e 16). A árvore filogenética do genoma completo (nucleotídeo) resultou em uma distinção maior entre CMaV e SYNV, tratando-os como sequências distintas.

O alinhamento de CMaV com os demais rhabdovírus resultou em uma identidade próxima ao SYNV a nível de aminoácido. Para o genoma completo (nucleotídeo), no entanto, a identidade (56,99%) não foi a mais alta entre os genomas representados. Os vírus do gênero *Nucleorhabdovirus* *Maize mosaic virus* (MMV) e *Maize Iranian mosaic virus* (MIMV) apresentaram uma identidade de 58,06%, enquanto que MMV e *Taro vein chlorosis virus* (TaVVCV) resultaram em 58,63%. Alguns vírus do gênero *Cytorhabdovirus* também apresentaram alta identidade para o genoma, como o *Lettuce yellow mottle virus* (LYMV) e *Lettuce necrotic yellows virus* (LNYV), com 59,27% de identidade, e os vírus *Northern cereal mosaic virus* (NCMV) e *Barley yellow striate mosaic virus* (BYSMV), com 55,75%. Para a proteína L, a identidade varia entre 21,60%, entre o NCMV e CMaV e 87,15%, entre SYNV e CMaV. Além de CMaV e SYNV, alguns vírus apresentaram alta identidade para a mesma proteína, como entre LNYV e LYMV, com 66,02%.

Para os rhabdovírus que infectam plantas, o critério para diferenciação entre espécies é baseado na definição da gama-de-hospedeiras e na especificidade com o inseto vetor. A análise das sequências do genoma obtidas para alguns vírus pertencentes ao gênero tem auxiliado na distinção entre espécies, porém, segundo o ICTV, este não é um critério suficiente para classificar espécies. Ainda segundo o ICTV, o sequenciamento de genomas, bem como a utilização de técnicas de RT-PCR e ensaios de localização de proteínas, podem ser ferramentas úteis para auxiliar a diferenciação entre espécies (Dietzgen et al., 2009).

Casos de espécies de nucleorhabdovírus com alta identidade a nível de aminoácido e nucleotídeo podem ser evidenciados pelo trabalho de Massah *et al.* (2008). Neste trabalho, verificou-se que o vírus MIMV apresenta maior identidade com os vírus MMV e TaVVCV. A proteína L, por exemplo, apresentou 64% de identidade de aminoácido entre MIMV e MMV (Massah *et al.*, 2008). Como não há critério de identidade para demarcação de espécies em rhabdovírus que infectam plantas, a justificativa para MIMV ser uma espécie nova foi com base na identidade a nível de aminoácido para a proteína L. Para os rhabdovírus que infectam animais *Vesicular stomatitis virus* (VSIV) e *Vesicular stomatitis New Jersey virus* (VSNJV), para espécies reconhecidamente distintas, a identidade para a proteína L é de 66%. Para os vírus serem considerados estirpes, a identidade sobe para 94,4%, como definido para VSNJV Ogd e VSNJV Haz, estirpes do mesmo vírus VSNJV (Dietzgen *et al.*, 2009).

A discriminação entre espécies de nucleorhabdovírus e cytorhabdovírus é feita, geralmente, com base na identidade para as proteínas L e N. Além disso, compara-se a filogenia com as demais espécies já definidas. Ainda assim, não há uma demarcação definida (Petrzik, 2012). Análises com a proteína L revelaram que no cytorhabdovírus *Strawberry crinkle virus* (SCV) havia dois subgrupos, considerados isolados, que apresentavam 98% de identidade com SCV. Os isolados, no entanto, apresentaram apenas 79% de identidade entre si (Klerks *et al.*, 2004). Para o nucleorhabdovírus TaVCoV, o gene L foi usado para comparar os isolados. As sequências nucleotídicas variaram até 27,4%, enquanto que a nível de aminoácido a diferença maior foi 11,3%. Assim, mesmo sem uma demarcação por identidade de sequências definida para a família, alguns autores consideram isolados os vírus que apresentarem a proteína L com identidade maior que 88,7% a nível de aminoácido (Petrzik, 2012).

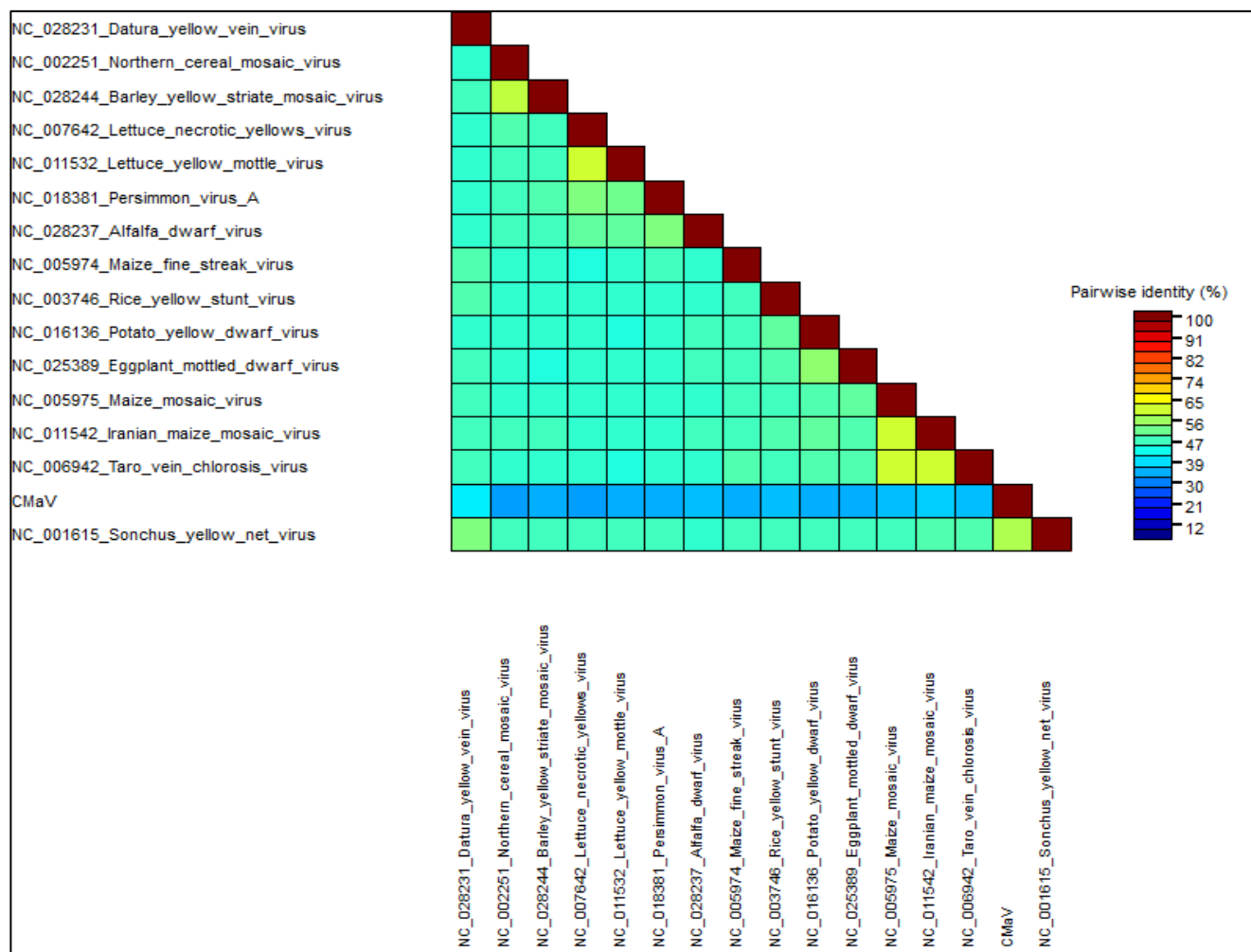


Figura 9. Matriz de porcentagem de identidade entre a sequência do genoma completo de *Coreopsis mottle-associated virus* (CMaV) obtida via sequenciamento Sanger de amostras de *Coreopsis lanceolata* com sequências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de nucleotídeo.

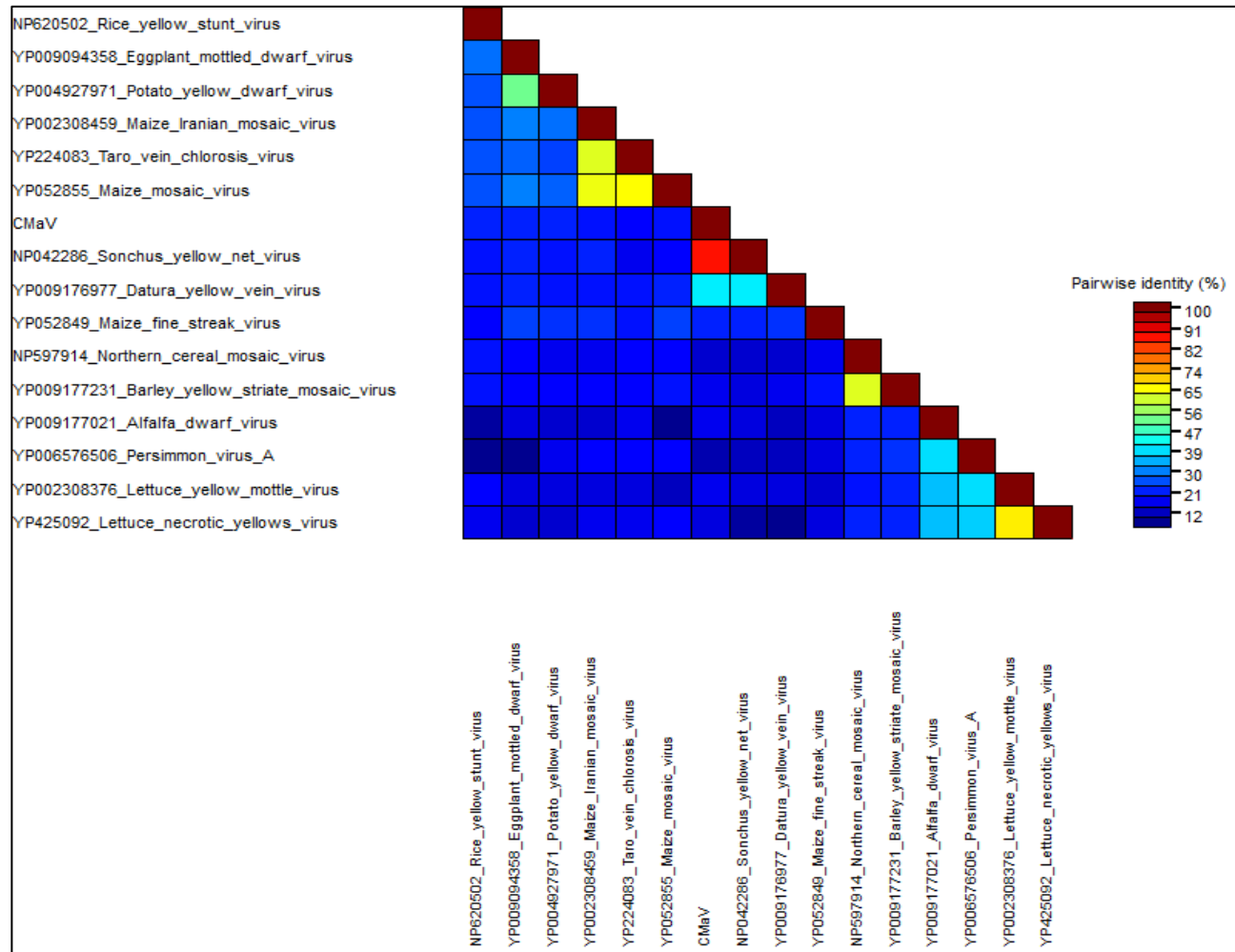


Figura 10. Matriz de porcentagem de identidade entre a sequência de aminoácidos da proteína L (polimerase) de *Coreopsis mottle-associated virus* (CMaV) obtida via sequenciamento Sanger de amostras de *Coreopsis lanceolata* e diferentes sequências de rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI).

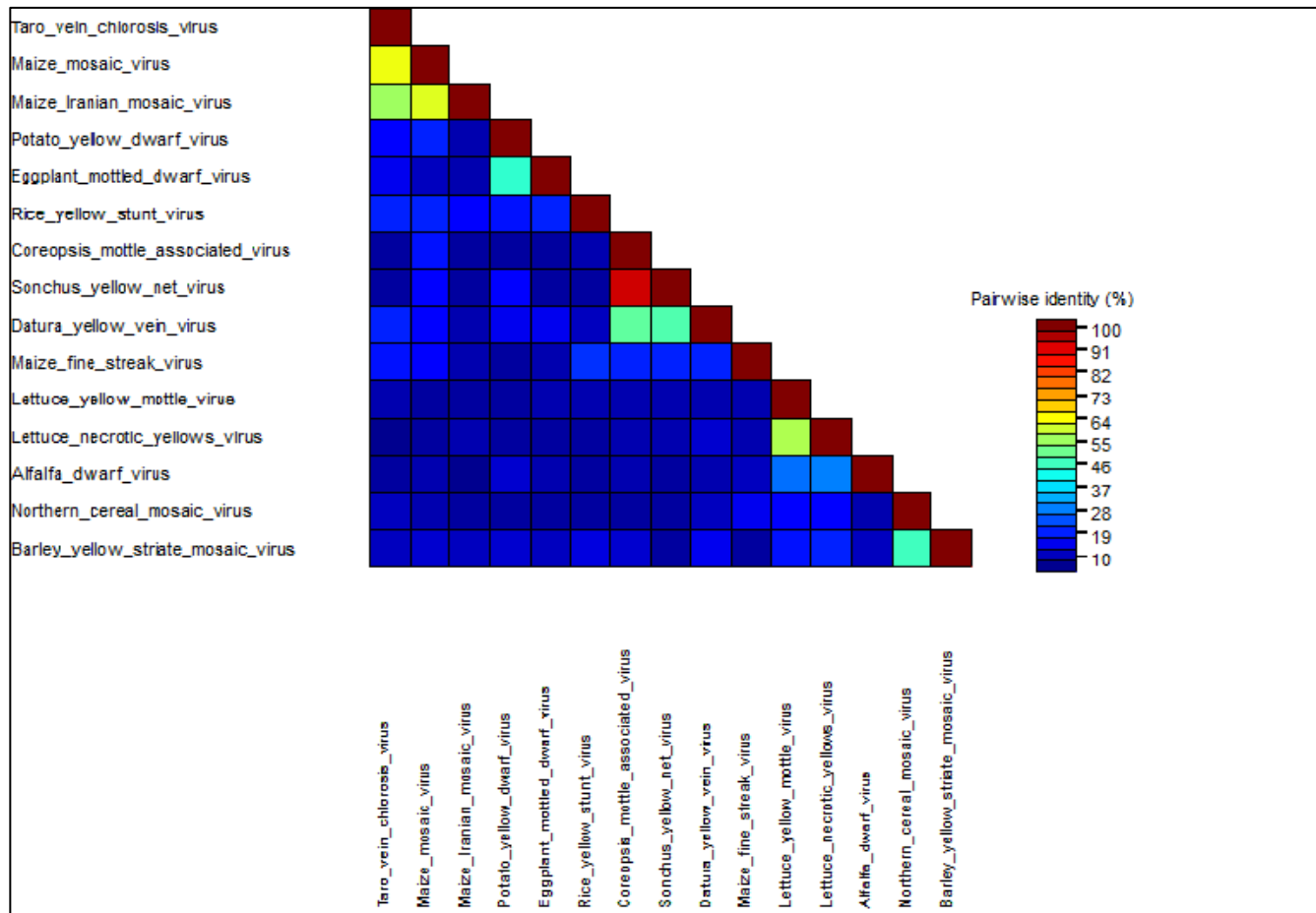


Figura 11. Matriz de porcentagem de identidade entre a sequência de aminoácidos da proteína N (nucleocapsídeo) de *Coreopsis mottle-associated virus* (CMaV) obtida via sequenciamento Sanger de amostras de *Coreopsis lanceolata* e diferentes sequências de rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI).

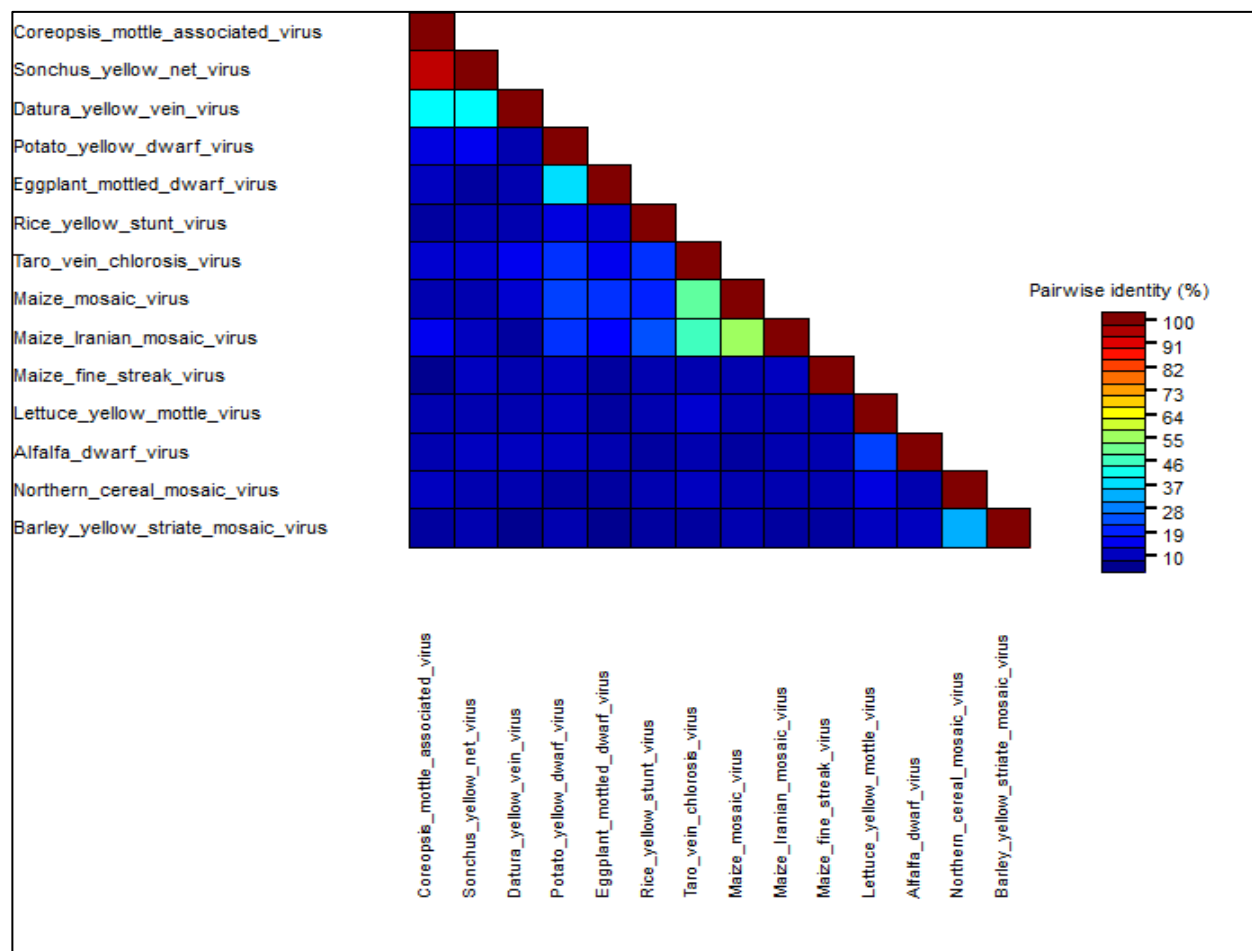


Figura 12. Matriz de porcentagem de identidade entre a sequência de aminoácidos da proteína G (glicoproteína) de *Coreopsis mottle-associated virus* (CMAV) obtida via sequenciamento Sanger de amostras de *Coreopsis lanceolata* e diferentes sequências de rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI).

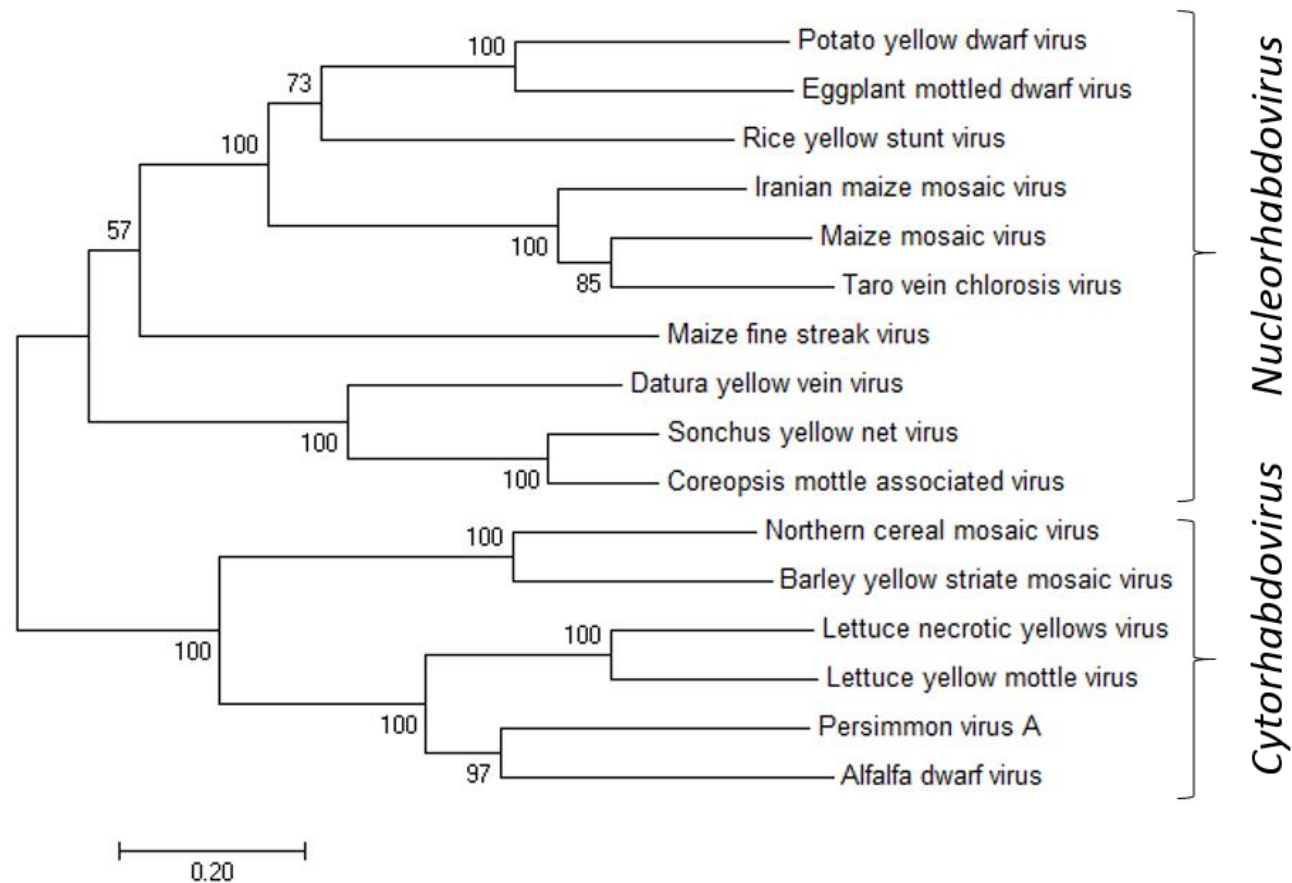


Figura 13. Árvore filogenética com base no algoritmo *Maximum Likelihood* (1000 repetições *bootstrap*) do genoma completo de *Coreopsis mottle-associated virus* (CMaV) obtida via sequenciamento Sanger de amostras de *Coreopsis lanceolata* com seqüências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de nucleotídeo. A espécie viral CMaV encontra-se no mesmo clado do *Sonchus yellow net nucleorhabdovirus*, indicando que faça parte do gênero *Nucleorhabdovirus*.

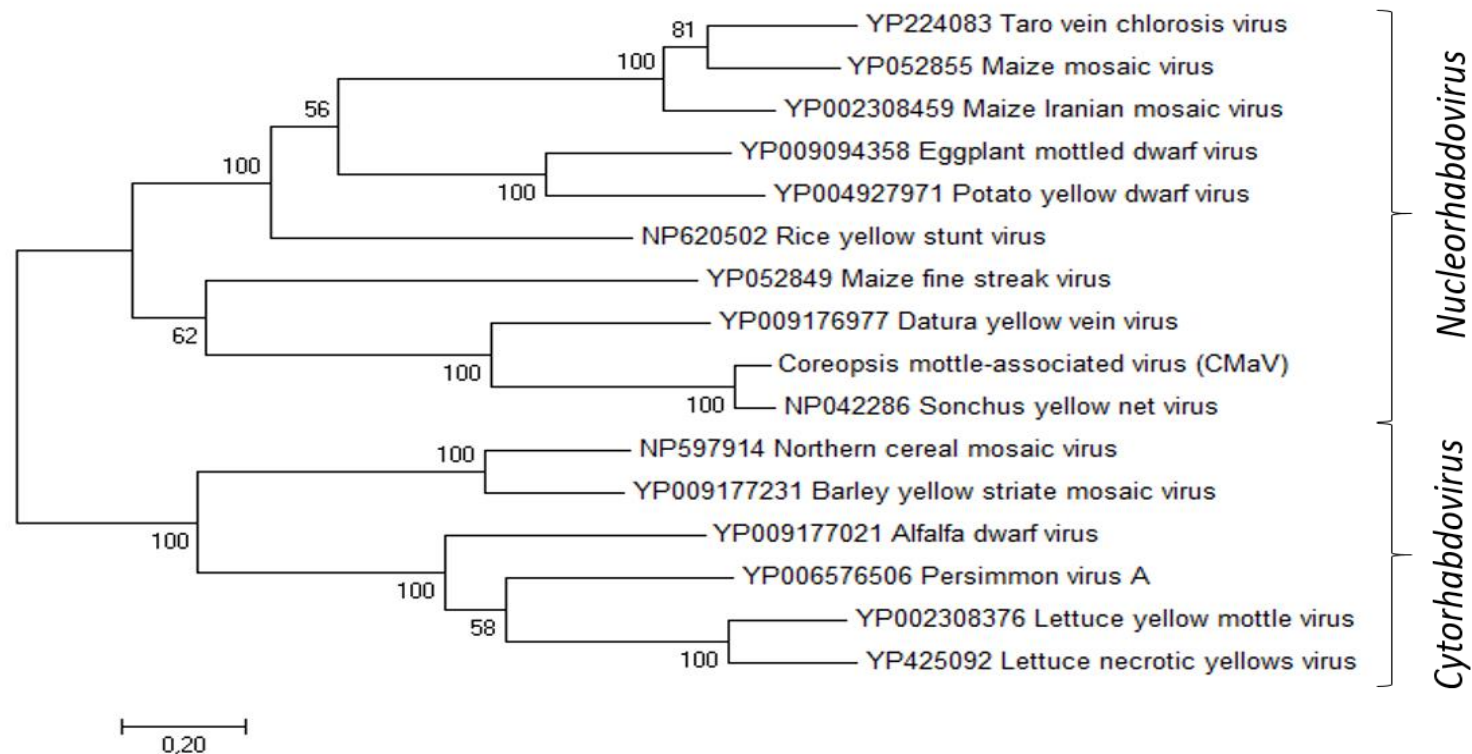


Figura 14. Árvore filogenética com base no algoritmo *Maximum Likelihood* (1000 repetições *bootstrap*) da proteína L (polimerase) do vírus *Coreopsis mottle-associated virus* (CMaV) obtida via sequenciamento Sanger de amostras de *Coreopsis lanceolata* com sequências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de aminoácido. A espécie viral CMaV encontra-se no mesmo clado do *Sonchus yellow net nucleorhabdovirus*, indicando que CMaV faça parte do gênero *Nucleorhabdovirus*.

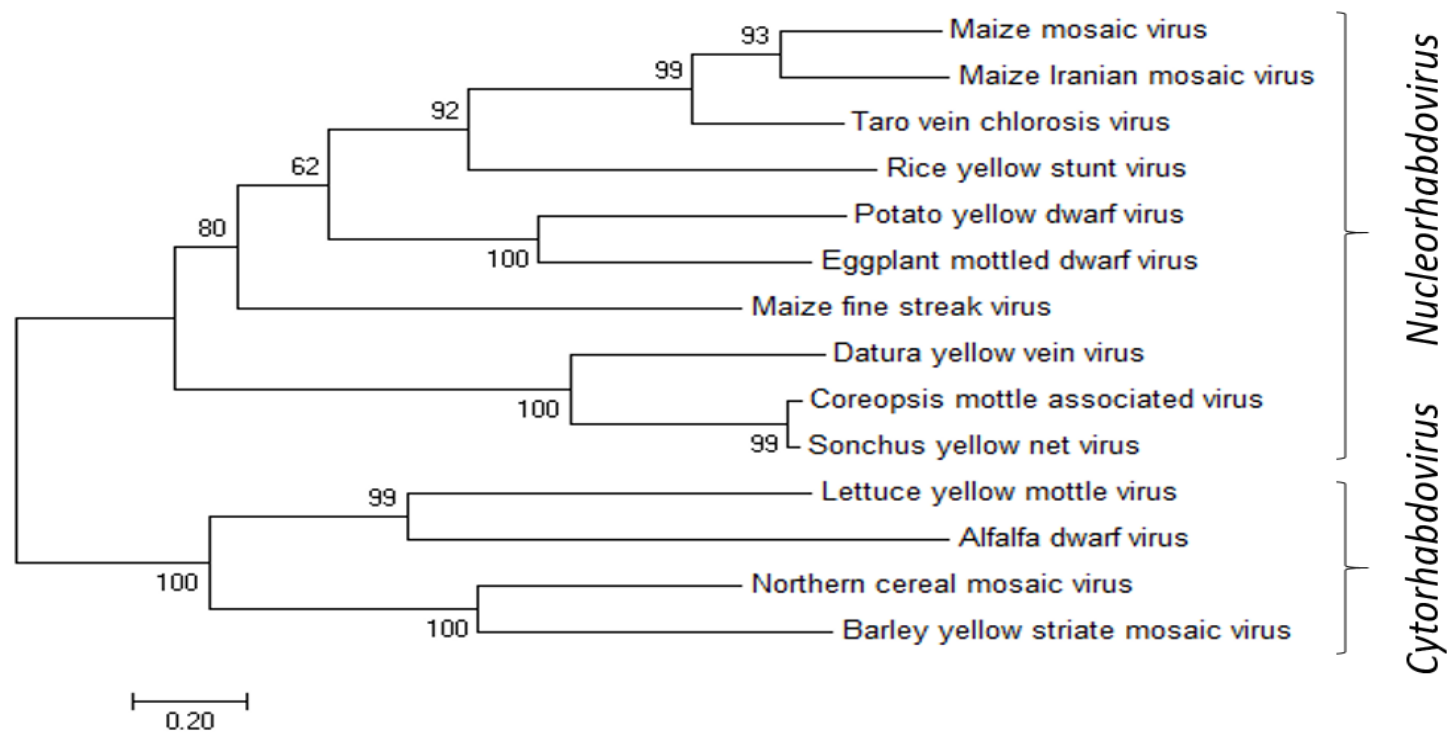


Figura 15. Árvore filogenética com base no algoritmo *Maximum Likelihood* (1000 repetições *bootstrap*) da proteína G (glicoproteína) do vírus *Coreopsis mottle-associated virus* (CMAV) obtida via sequenciamento Sanger de amostras de *Coreopsis lanceolata* com sequências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de aminoácido. A espécie viral CMAV encontra-se no mesmo clado do *Sonchus yellow net nucleorhabdovirus*, indicando que CMAV faça parte do gênero *Nucleorhabdovirus*.

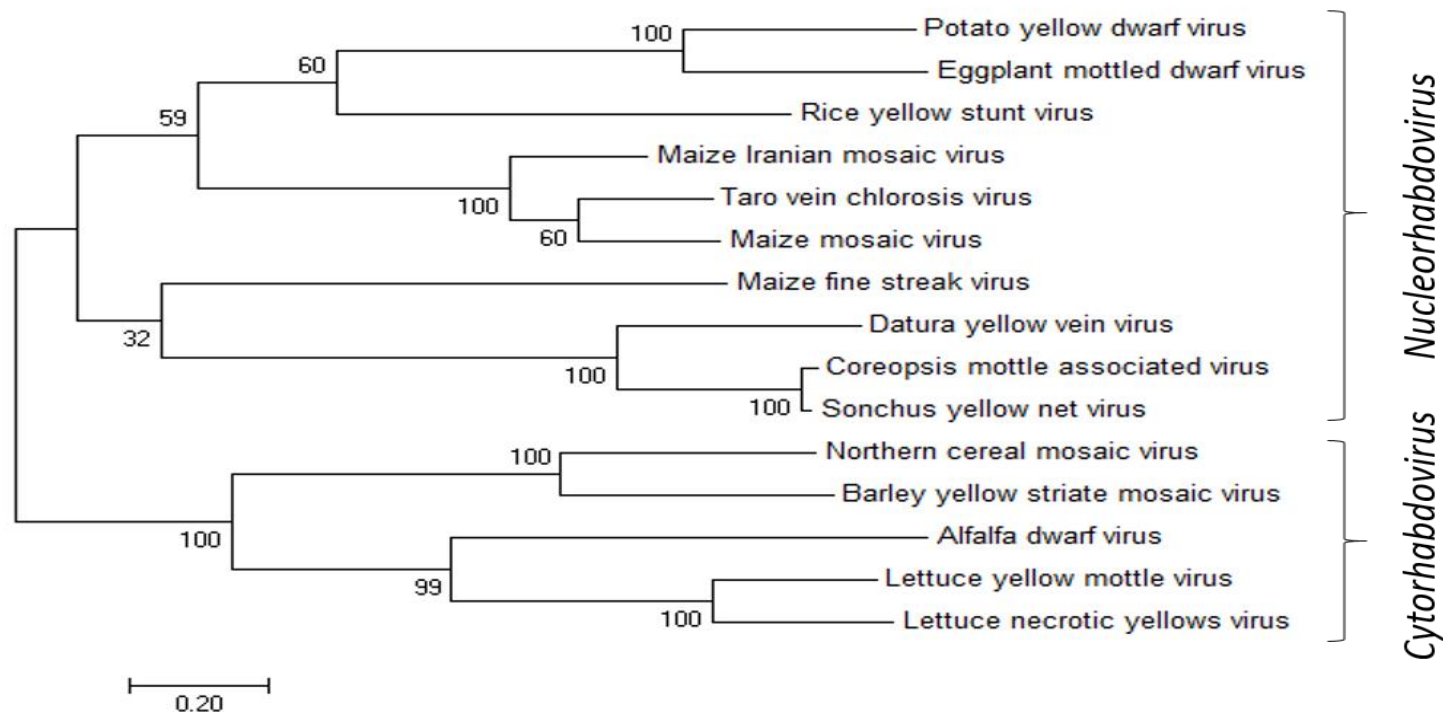


Figura 16. Árvore filogenética com base no algoritmo *Maximum Likelihood* (1000 repetições *bootstrap*) da proteína N (nucleocapsídeo) do vírus *Coreopsis mottle-associated virus* (CMaV) obtida via sequenciamento Sanger de amostras de *Coreopsis lanceolata* com sequências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de aminoácido. A espécie viral CMaV encontra-se no mesmo clado do *Sonchus yellow net nucleorhabdovirus*, indicando que CMaV faça parte do gênero *Nucleorhabdovirus*.

A análise da sequência do genoma de PMaV comparado aos demais rhabdovírus por alinhamento múltiplo de sequências e *pairwise* mostrou sua maior identidade com o PYDV e sua relação próxima ao gênero *Nucleorhabdovirus*. A nível de nucleotídeo, o genoma de PMaV apresentou 46,99% de identidade ao genoma de PYDV, seguido de *Eggplant mottled dwarf virus* (EMDV), com 37,31% de identidade. Os vírus EMDV e PYDV apresentaram 51,88% de identidade, ou seja, a nível de nucleotídeo o PYDV é mais próximo de EMDV do que de PMaV (**Figura 17**).

Em nível de aminoácido, o alinhamento para as proteínas N, G e L também mostrou a relação próxima de PMaV com PYDV e EMDV. A proteína N (nucleocapsídeo) apresentou 67,74% de identidade entre PMaV e PYDV e 47,17% de identidade entre PMaV e EMDV. O vírus PYDV apresentou 46,77% de identidade com EMDV (**Figura 19**). A proteína G (glicoproteína) resultou em 64,68% de identidade entre PMaV e PYDV e 39,93% entre PMaV e EMDV. Entre os vírus PYDV e EMDV foi observado 38,27% de identidade (**Figura 20**). Para a proteína L (polimerase), a identidade entre PMaV e PYDV foi de 70,91%, enquanto que entre PYDV e EMDV, 52,34% (**Figura 18**). O alinhamento de aminoácido para as proteínas N, G e L, portanto, mostraram que PYDV é mais próximo de PMaV, quando comparado ao EMDV.

A comparação por alinhamento *pairwise* entre as proteínas X e Y de PMaV e PYDV resultou em uma identidade de 13,9% para a proteína X e 67,3% para a proteína Y (proteína de movimento). Para a proteína X, os *gaps* representaram 46,3%, enquanto que em Y, 8,7%.

As árvores filogenéticas do genoma completo (a nível de nucleotídeo) (**Figura 21**) e das proteínas N, G e L (aminoácido) (**Figuras 22, 23 e 24**) mostraram a relação próxima de PMaV com PYDV, além de apresentar um grupo monofilético composto por EMDV, PYDV e PMaV. As árvores também evidenciam que PMaV pertence ao gênero *Nucleorhabdovirus*, juntamente com os mesmos representados.

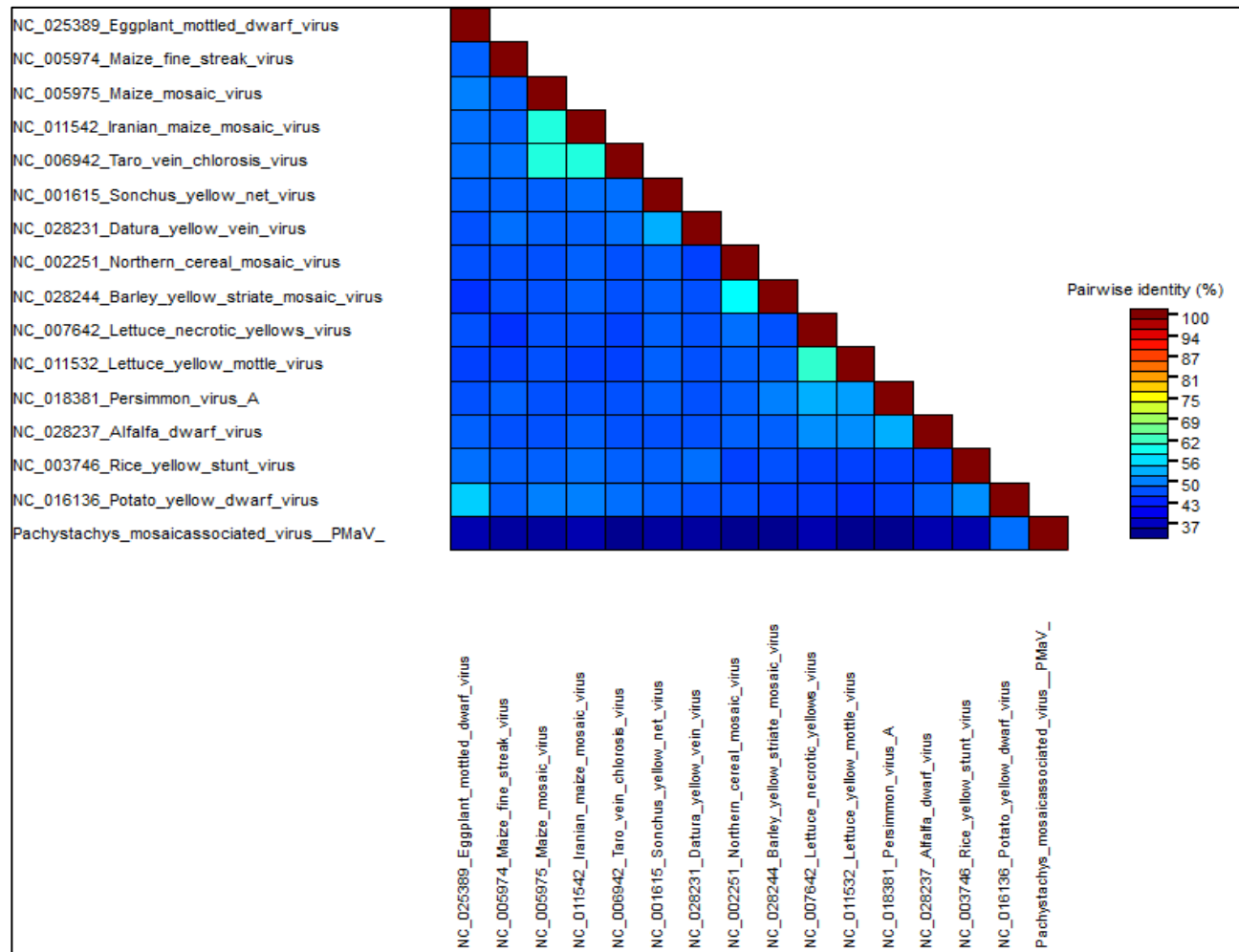


Figura 17. Matriz de porcentagem de identidade entre a sequência do genoma completo de Pachystachys mosaica-associated virus (PMAV) obtida via sequenciamento Sanger de amostras de *Pachystachys lutea* com sequências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de nucleotídeo.

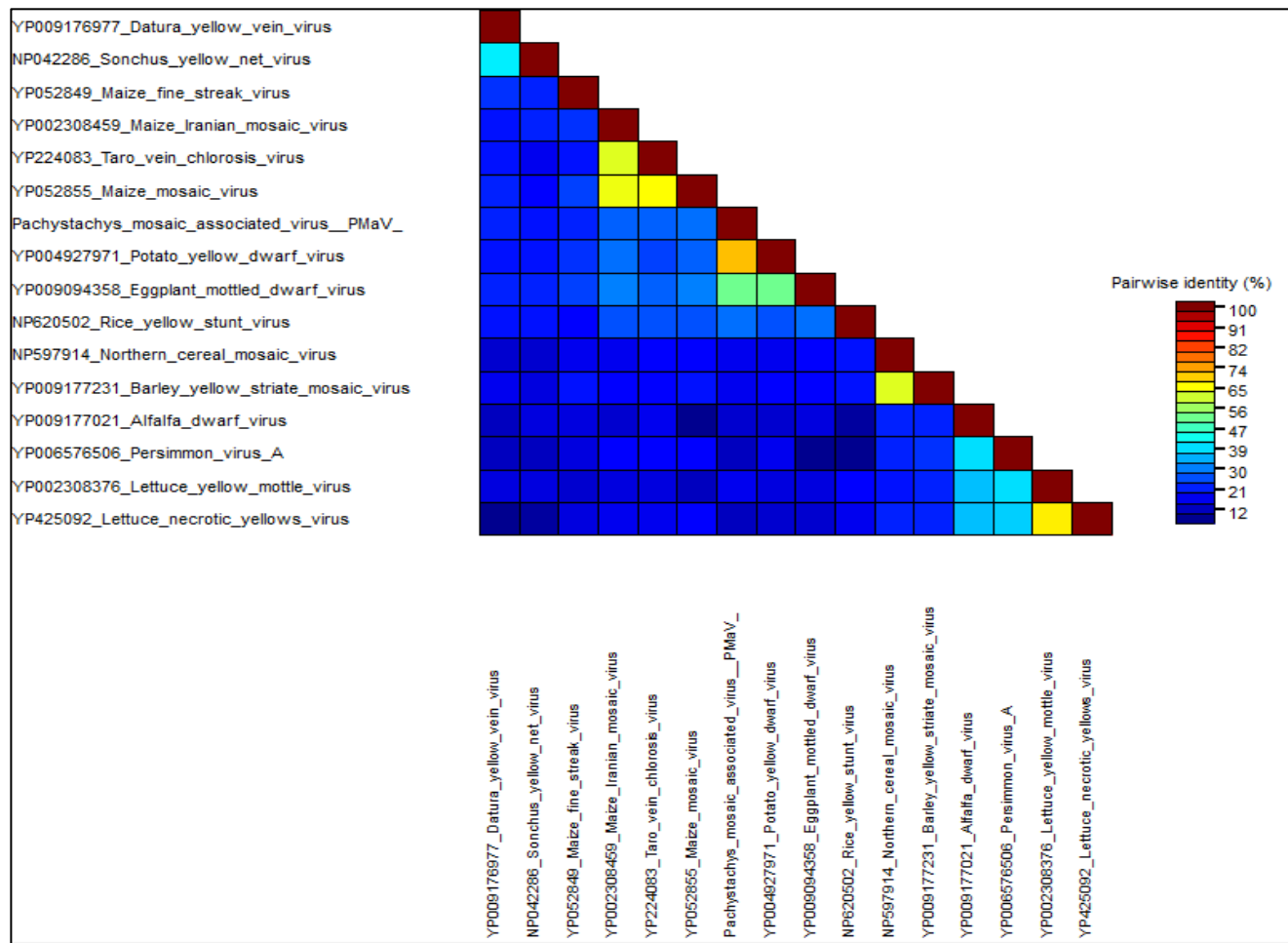


Figura 18. Matriz de porcentagem de identidade entre a sequência de aminoácidos da proteína L (polimerase) de Pachystachys mosaic-associated virus (PMAV) obtida via sequenciamento Sanger de amostras de *Pachystachys lutea* e diferentes sequências de rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI).

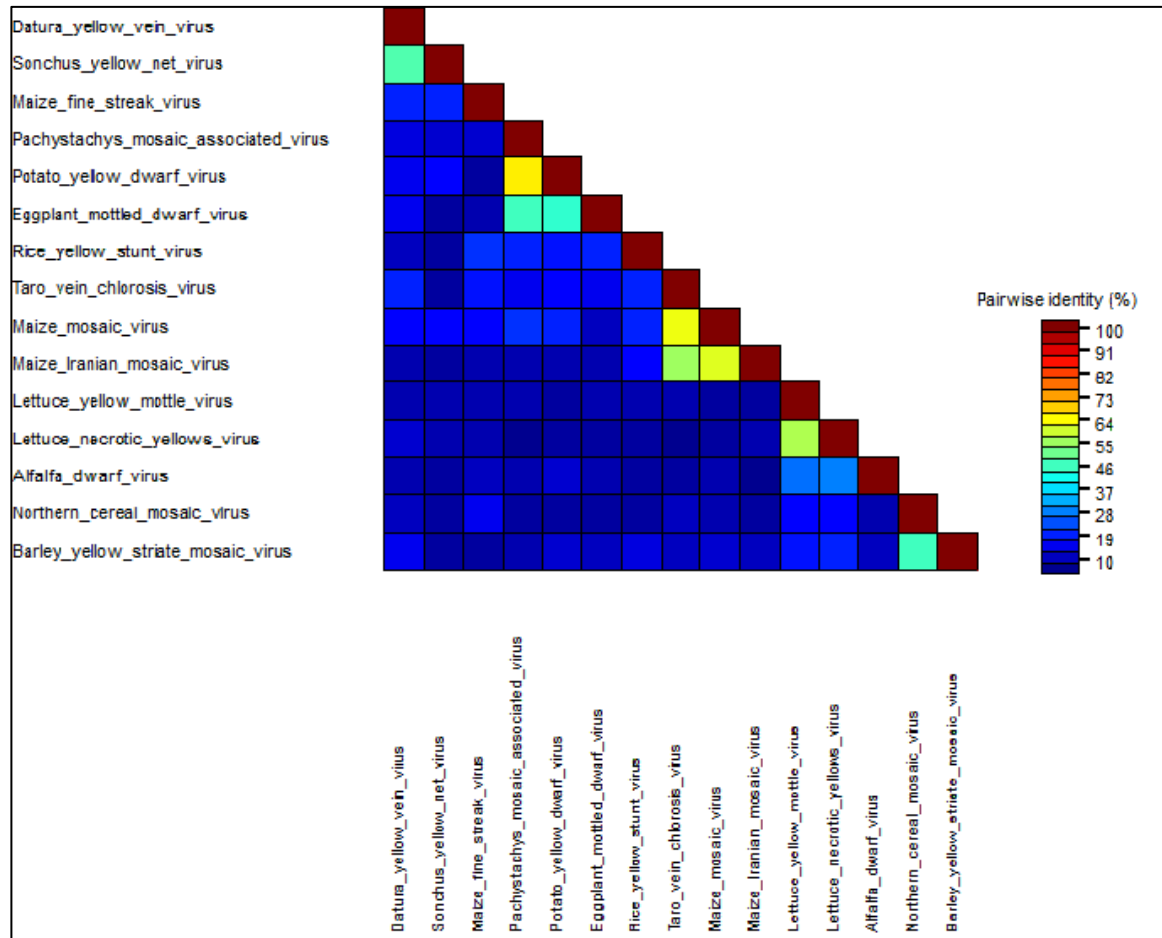


Figura 19. Matriz de porcentagem de identidade entre a sequência de aminoácidos da proteína N (nucleocapsídeo) de *Pachystachys mosaic-associated virus* (PMAV) obtida via sequenciamento Sanger de amostras de *Pachystachys lutea* e diferentes sequências de rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI).

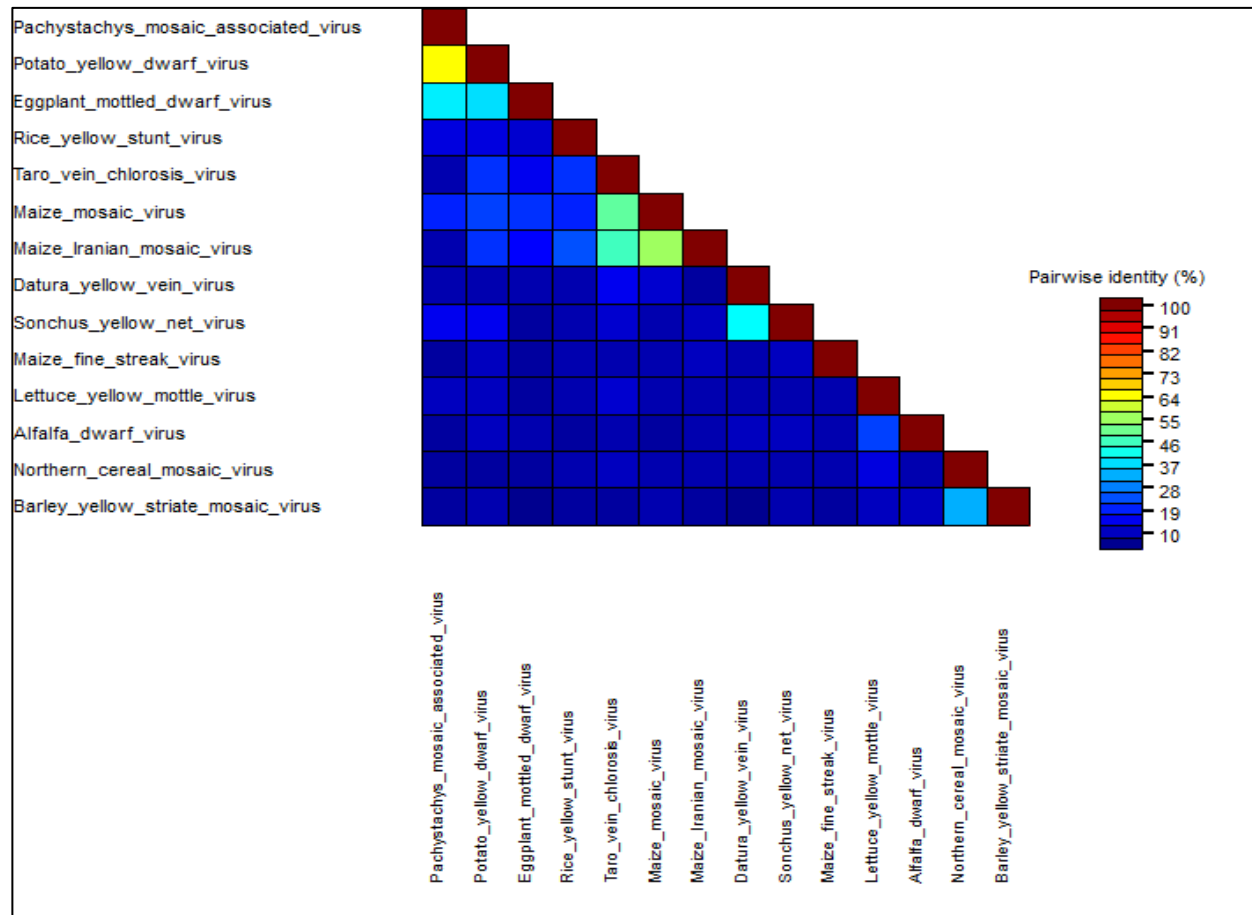


Figura 20. Matriz de porcentagem de identidade entre a sequência de aminoácidos da proteína G (glicoproteína) de Pachystachys mosaic-associated virus (PMAV) obtida via sequenciamento Sanger de amostras de *Pachystachys lutea* e diferentes sequências de rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI).

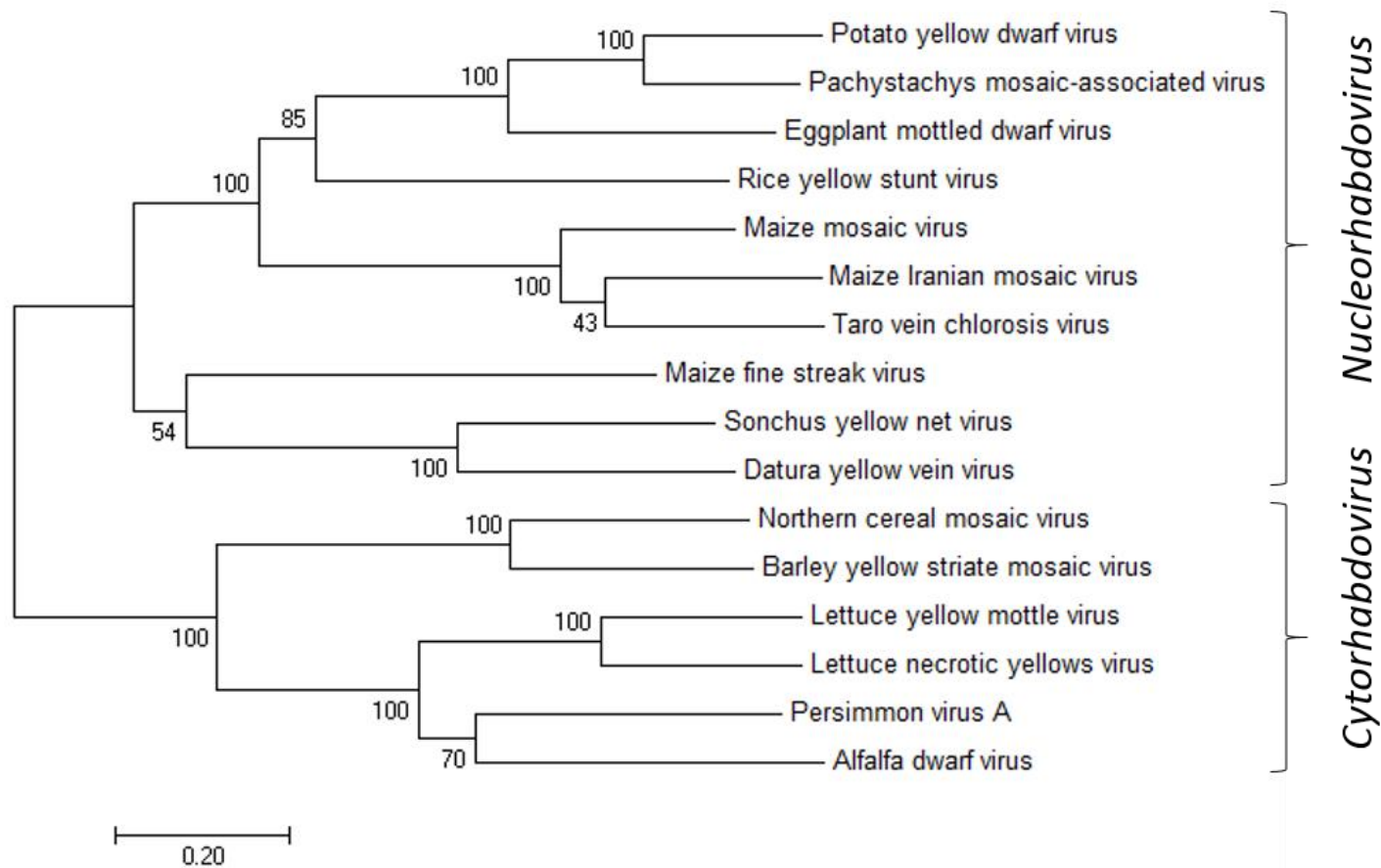


Figura 21. Árvore filogenética com base no algoritmo *Maximum Likelihood* (1000 repetições *bootstrap*) do genoma completo de *Pachystachys mosaic-associated virus* (PMAV) obtida via sequenciamento Sanger de amostras de *Pachystachys lutea* com sequências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de nucleotídeo. A espécie viral PMAV encontra-se no mesmo clado do *Potato yellow dwarf nucleorhabdovirus* (PYDV) e fazendo parte do gênero *Nucleorhabdovirus*.

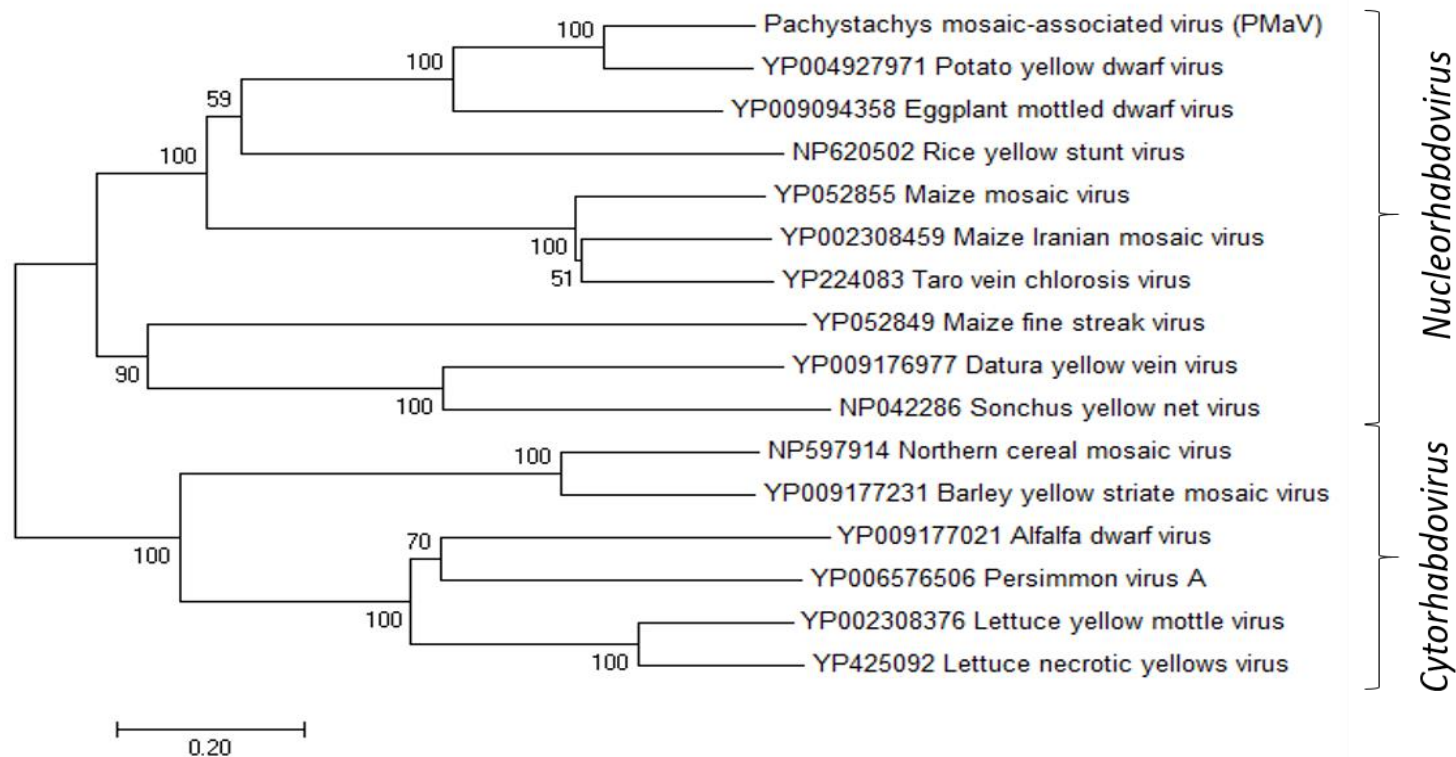


Figura 22. Árvore filogenética com base no algoritmo *Maximum Likelihood* (1000 repetições *bootstrap*) da proteína L (polimerase) de *Pachystachys mosaic-associated virus* (PMAV) obtida via sequenciamento Sanger de amostras de *Pachystachys lutea* com seqüências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de aminoácido. A espécie viral PMAV encontra-se no mesmo clado do *Potato yellow dwarf nucleorhabdovirus* (PYDV) e fazendo parte do gênero *Nucleorhabdovirus*.

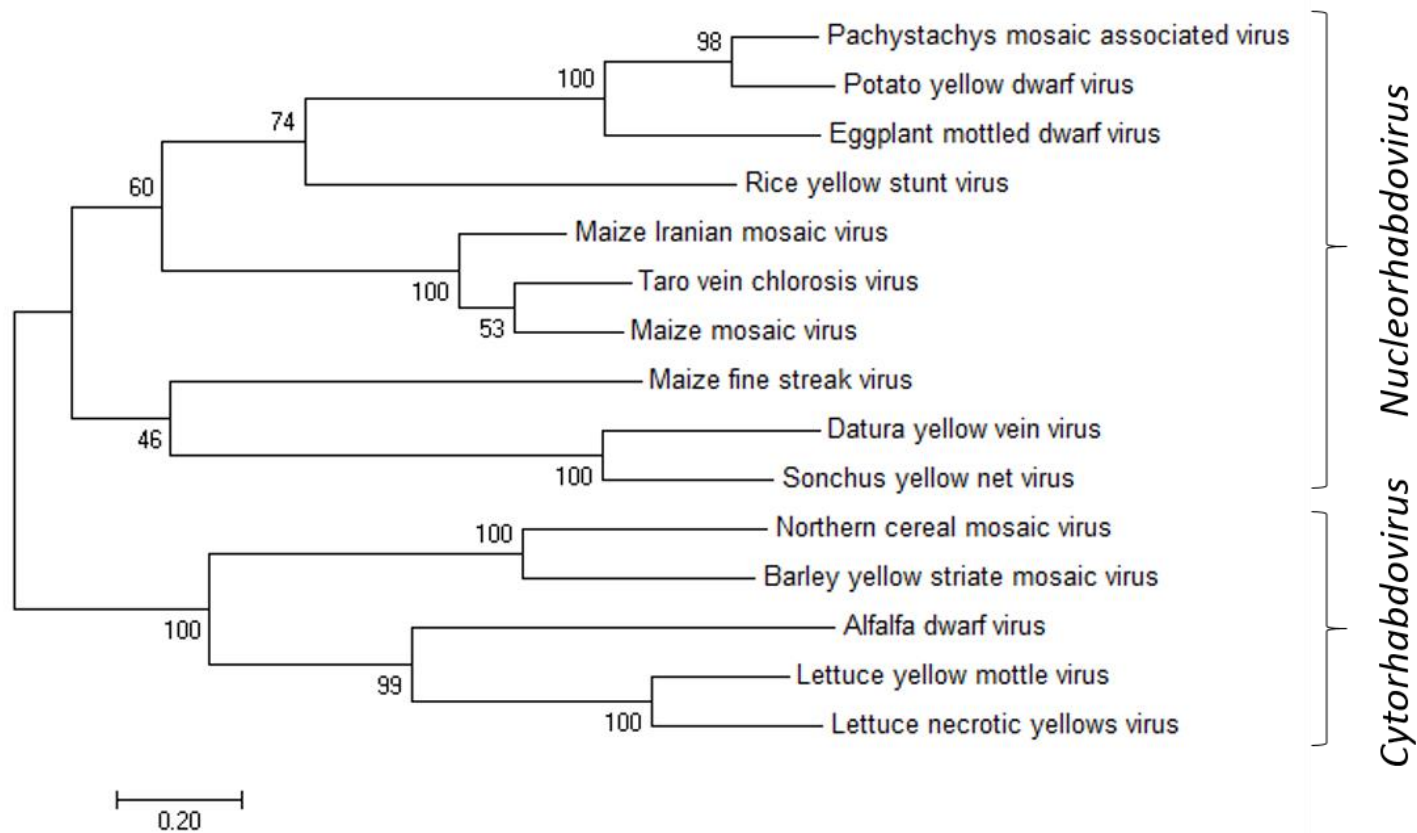


Figura 23. Árvore filogenética com base no algoritmo *Maximum Likelihood* (1000 repetições *bootstrap*) da proteína N (nucleocapsídeo) de *Pachystachys mosaic-associated virus* (PMaV) obtida via sequenciamento Sanger de amostras de *Pachystachys lutea* com seqüências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de aminoácido. A espécie viral PMaV encontra-se no mesmo clado do *Potato yellow dwarf nucleorhabdovirus* (PYDV) e fazendo parte do gênero *Nucleorhabdovirus*.

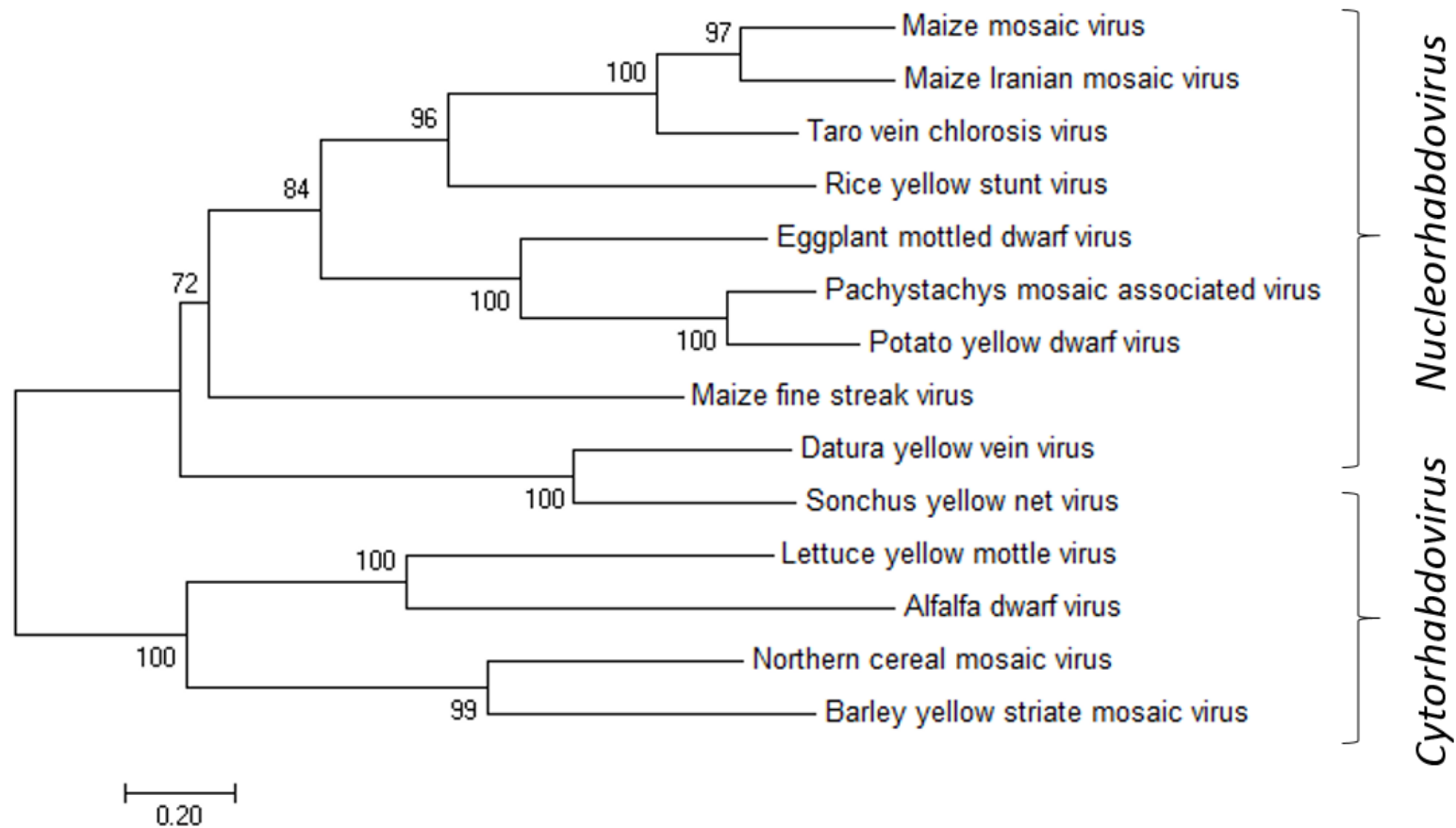


Figura 24. Árvore filogenética com base no algoritmo *Maximum Likelihood* (1000 repetições *bootstrap*) da proteína G (glicoproteína) de *Pachystachys mosaic-associated virus* (PMAV) obtida via sequenciamento Sanger de amostras de *Pachystachys lutea* com sequências de diferentes rhabdovírus obtidas diretamente do *National Center for Biotechnology Information* (NCBI) a nível de aminoácido. A espécie viral PMAV encontra-se no mesmo clado do *Potato yellow dwarf nucleorhabdovirus* (PYDV) e fazendo parte do gênero *Nucleorhabdovirus*.

Para quantificar a real distinção genômica entre os vírus recuperados e seus genomas de referência, usou-se diferentes metodologias e ferramentas filogenéticas. No caso das espécies virais CMaV e PMaV, após o alinhamento múltiplo de sequências (MSA, do inglês *Multiple Sequence Alignment*) a árvore filogenética foi obtida. O MSA distingue diferentes sequências, concebendo a identidade (em %) entre elas. Como a família *Rhabdoviridae* não apresenta um critério de identidade fixo e geral, utiliza-se outros critérios, como hospedeiros e estrutura do genoma. Com o alinhamento, é possível, também, realizar árvores filogenéticas para análises evolutivas e, às vezes, filogeográficas das sequências alinhadas.

4. CONCLUSÕES GERAIS E PERSPECTIVAS

A tecnologia do *Next Generation Sequencing* (NGS) confirmou ser uma poderosa ferramenta para a análise da diversidade viral em plantas pouco estudadas quanto a infecção e/ou presença de vírus, como as plantas ornamentais. Diferentes *softwares* de bioinformática podem explorar os resultados de um mesmo sequenciamento NGS, inclusive alterando parâmetros de um mesmo programa, como o *k-mer* no montador *Velvet*. Os dados gerados por diferentes programas auxiliam na montagem de sequências de genomas virais, além da interpretação das informações geradas a partir do sequenciamento como um todo. A partir do sequenciamento NGS, notou-se a presença de possíveis sequências de vírus pertencentes às famílias virais *Tombusviridae*, *Rhabdoviridae* e *Potyviridae*.

Neste trabalho foram identificadas, sequenciadas e recuperadas molecularmente duas prováveis espécies novas de *Nucleorhabdovirus* (*Rhabdoviridae*), *Coreopsis mottle-associated virus* (CMaV) e *Pachystachys mosaic-associated virus* (PMaV), infectando as plantas ornamentais *Coreopsis lanceolata* (margaridinha) e *Pachystachys lutea* (camarão-amarelo), respectivamente. Além desses, outras quatro sequências virais foram detectadas no *pool* das dez amostras de plantas ornamentais, um *Umbravirus* (*Tombusviridae*), um *Pelarspovirus* (*Tombusviridae*), um *Potyvirus* (*Potyviridae*) e um *Cytorhabdovirus* (*Rhabdoviridae*).

As espécies CMaV e PMaV apresentaram como genomas de referência SYNIV e PYDV, respectivamente. O processo de detecção abriu caminho para a recuperação completa do genoma de ambos os vírus. O resequenciamento dos vírus via Sanger é uma maneira de assegurar e comprovar que o genoma montado *in silico* era, de fato, real. Utilizando PCR comum para amplificar as diferentes regiões do genoma e o protocolo 3' e 5' RACE para as regiões terminais, foi possível resgatar todo o genoma dos dois vírus. Com o sequenciamento Sanger, viu-se que pouco mudou em relação a sequência de NGS, validando todo o processo

de montagem e bioinformática. Como as novas sequências se mostraram diferentes, além de pertencerem a hospedeiros distintos das usadas como referência, acusou-se que são espécies novas do gênero *Nucleorhabdovirus*.

A gama de grupos virais presentes no NGS mostrou o potencial de inóculo das plantas ornamentais, tipicamente negligenciadas em estudos fitopatológicos. Viu-se que nesse grupo de hospedeiras encontra-se uma série de espécies ainda a serem caracterizadas. Há, portanto, a possibilidade de alguma espécie nova encontrada em plantas ornamentais poder ocasionar prejuízos em outras culturas. Para isso, portanto, pesquisas devem ser direcionadas à ensaios de transmissão, bem como de gama-de-hospedeiros dos vírus já caracterizados e avaliar as demais sequências presentes nos dados de NGS.

Para definir a interação patógeno-hospedeiro e cumprir os postulados de Koch, os vírus CMaV e PMaV precisam ser estudados em mais detalhes, incluindo a caracterização biológica. A definição de sua natureza não só auxiliará em um melhor entendimento acerca de vírus em ornamentais, como para o grupo de vírus de plantas para a família em geral.

5. REFERÊNCIAS BIBLIOGRÁFICAS

Adams IP, Glover RH, Monger WA, *et al.*, 2009a. Next-Generation Sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Molecular Plant Pathology* **10**, 537-45.

Adams IP, Miano DW, Kinyua ZM, *et al.*, 2013. Use of Next-Generation Sequencing for the identification and characterization of Maize chlorotic mottle virus and Sugarcane mosaic virus causing maize lethal necrosis in Kenya. *Plant Pathology* **62**, 8.

Adams MJ, Antoniw JF, Beaudoin F, 2005a. Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. *Molecular Plant Pathology* **6**, 471-87.

Adams MJ, Antoniw JF, Fauquet CM, 2005b. Molecular criteria for genus and species discrimination within the family Potyviridae. *Archives of Virology* **150**, 459-79.

Adams MJ, Zerbini FM, French R, Rabenstein F, Stenger DC, Valkonen JPT, 2009b. *Virus Taxonomy: Ninth Report - Potyviridae family*. ICTV.

Al Rwahnih M, Daubert S, Golino D, Rowhani A, 2009. Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology* **387**, 395-401.

Alexandre MaV, Duarte LML, Rivas EB, Galleti SR, 2010. Vírus detectados em plantas ornamentais no período 2004 a 2008. *Revista Brasileira de Horticultura Ornamental* **16**, 95-100.

Alonso AM, Silva JCS, 2009. Floricultura no Distrito Federal: perspectivas para o fortalecimento da cadeia produtiva de flores e plantas ornamentais. In. Planaltina, DF: Embrapa Cerrados. (2015.)

Ansorge WJ, 2009. Next-generation DNA sequencing techniques. *New Biotechnology* **25**, 195-203.

Baker KF, Linderman RG, 1979. Unique features of the pathology of ornamental plants. *Annual Review Phytopathology* **17**, 25.

Bankevich A, Nurk S, Antipov D, *et al.*, 2012. *SPAdes*: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**, 455-77.

Bao S, Jiang R, Kwan W, Wang B, Ma X, Song YQ, 2011. Evaluation of *Next-Generation Sequencing* software in mapping and assembly. *Journal of Human Genetics* **56**, 406-14.

Barnett OW, 1991. Potyviridae, a proposed family of plant viruses. *Archives of Virology* **118**, 139-41.

Bolger AM, Lohse M, Usadel B, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20.

Brown JK, Fauquet CM, Briddon RW, Zerbini M, Moriones E, Navas-Castillo J, 2009. *Virus Taxonomy: Ninth Report - Geminiviridae family*. ICTV.

Chen J, Chen J, Adams MJ, 2001. A universal PCR *primer* to detect members of the Potyviridae and its use to examine the taxonomic status of several members of the family. *Archives of Virology* **146**, 757-66.

De Stradis A, Parrella G, Vovlas C, Ragozzino A, 2008. Vein yellowing of *Hibiscus rosa-sinensis* caused by Eggplant mottled dwarf virus in southern Italy. *Journal of Plant Pathology* **90**, 3.

Dietzgen RG, Calisher CH, Kurath G, *et al.*, 2009. *Virus Taxonomy: Ninth Report - Rhabdoviridae family*. ICTV.

Dietzgen RG, Kuhn JH, Clawson AN, *et al.*, 2014. Dichorhavirus: a proposed new genus for Brevipalpus mite-transmitted, nuclear, bacilliform, bipartite, negative-strand RNA plant viruses. *Archives of Virology* **159**, 607-19.

Filho Fde A, Nicolini C, Resende Rde O, Andrade GP, Pio-Ribeiro G, Nagata T, 2013. The complete genome sequence of a Brazilian isolate of yam mild mosaic virus. *Archives of Virology* **158**, 515-8.

Freitas-Astua J, Rodrigues JC, Kitajima EW, 2010. An annotated list of ornamentals naturally found infected by Brevipalpus mite-transmitted viruses. *Scientia Agricola* **67**, 348-71.

Gibbs MJ, Ziegler A, Robinson DJ, Waterhouse PM, Cooper A, 1996. Carrot mottle mimic virus (CMoMV): A second umbravirus associated with carrot motley dwarf disease recognised by nucleic acid hybridisation. In. <http://www.bspp.org.uk/mppol/1996/1111gibbs>: *Molecular Plant Pathology On-Line*. (2017.)

Goodin MM, Austin J, Tobias R, Fujita M, Morales C, Jackson AO, 2001. Interactions and nuclear import of the N and P proteins of sonchus yellow net virus, a plant *Nucleorhabdovirus*. *Journal of Virology* **75**, 9393-406.

Harakava R, Rivas EB, Alexandre MaV, Toscano AN, Duarte LML, 2008. Identificação e controle do Alternanthera mosaic virus isolado de Torenia sp. (Scrophulariaceae) *Revista Brasileira de Horticultura Ornamental* **14**, 59-66.

Higgins CM, Bejerman N, Li M, *et al.*, 2016. Complete genome sequence of Colocasia bobone disease-associated virus, a putative cytorhabdovirus infecting taro. *Archives of Virology* **161**, 745-8.

Ho T, Tzanetakis IE, 2014. Development of a virus detection and discovery pipeline using next generation sequencing. *Virology* **471-473**, 54-60.

Hull R, 2013. *Plant Virology*. Elsevier.

Ishikawa EA, Blawid R, Melo FL, *et al.*, 2016. Complete genome sequence of a putative new secovirus infecting yam (*Dioscorea*) plants. *Archives of Virology* **162**, 1.

Jackson AO, Dietzgen RG, Goodin MM, Bragg JN, Deng M, 2005. Biology of plant rhabdoviruses. *Annual Review Phytopathology* **43**, 623-60.

Junqueira AH, Peetz MS, 2015. Flores e Plantas ornamentais do Brasil - Série estudos mercadológicos. In: Sebrae SBDaÀMEPE-, ed. Brasília - DF: SEBRAE, 1-44. (1.)

Kaur C, Kumar S, Snehi SK, Raj SK, 2013. Molecular detection of Jasmine potyvirus associated with yellow mosaic symptoms on *Jasminum sambac* L. in India. *Plant Molecular Virology* **46**, 8.

Kearse M, Moir R, Wilson A, *et al.*, 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-9.

Kehoe MA, Coutts BA, Buirchell BJ, Jones RA, 2014. Plant virology and next generation sequencing: experiences with a Potyvirus. *PLoS One* **9**, e104580.

King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, 2011. *Ninth Report of the International Committee on Taxonomy of Viruses: Classification and Nomenclature of Viruses*. Elsevier.
Klerks MM, Lindner JL, Vaskova D, *et al.*, 2004. Detection and tentative grouping of Strawberry crinkle virus isolates. *European Journal of Plant Pathology* **110**, 7.

Lara-Victoriano F, Castillo-Reyes F, Flores-Gallegos C, Aguilar CN, Herrera-Rodríguez R, 2011. Metagenomics in plant pathology. *Phytopathology in the Omics Era* **2**, 9.

Marz M, Beerenwinkel N, Drosten C, *et al.*, 2014. Challenges in RNA virus bioinformatics. *Bioinformatics* **30**, 1793-9.

Massah A, Izadpanah K, Afsharifar AR, Winter S, 2008. Analysis of nucleotide sequence of Iranian maize mosaic virus confirms its identity as a distinct *Nucleorhabdovirus*. *Archives of Virology* **153**, 1041-7.

Matz M, Shagin D, Bogdanova E, *et al.*, 1999. Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Research* **27**, 1558-60.

Mcdonald JG, Singh RP, 1996. Host Range, Symptomology, and Serology of Isolates of Potato virus Y (PVY) That Share Properties with Both the PVYn and PVYo Strain Groups. *American Potato Journal* **73**, 7.

Mokili JL, Rohwer F, Dutilh BE, 2012. Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology* **2**, 63-77.

Mollov D, Lockhart B, Zlesak DC, 2013. Complete nucleotide sequence of Rosa rugosa leaf distortion virus, a new member of the family Tombusviridae. *Archives of Virology* **158**, 2617-20.

Muhire BM, Varsani A, Martin DP, 2014. SDT: a virus classification tool based on *pairwise* sequence alignment and identity calculation. *PLoS One* **9**, e108277.

Petrzik K, 2012. Bioinformatic Analysis of the L polymerase Gene Leads to Discrimination of New Rhabdoviruses. *Journal Phytopathology* **160**, 4.

Pop M, Salzberg SL, 2008. Bioinformatics challenges of new sequencing technology. *Trends in Genetics* **24**, 142-9.

Prabha K, Baranwal VK, Jain RK, 2013. Applications of next generation high throughput sequencing technologies in characterization, discovery and molecular interaction of plant viruses. *Indian Journal of Virology* **24**, 157-65.

Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N, 2012. Application of *Next-Generation Sequencing* technologies in virology. *Journal of General Virology* **93**, 1853-68.

Reeder GS, Knudson DL, Macleod R, 1972. The ribonucleic acid of potato yellow dwarf virus. *Virology* **50**, 301-4.

Rivas EB, Duarte LML, Alexandre MaV, Fernandes FMC, Harakava R, Chagas CM, 2005. A new Badnavirus species detected in Bougainvillea in Brazil. *Journal of General Plant Pathology* **71**, 3.

Rivas EB, Galleti SR, Duarte LML, Alexandre MaV, Cilli A, Estelita MEM, 2003. Detecção de Potyviridae em espécies de Dieffenbachia. *Arquivos do Instituto Biológico* **70**, 3.

- Rochon D, Lommel S, Martelli GP, Rubino L, Russo M, 2009. *Virus Taxonomy: Ninth Report - Tombusviridae family*. ICTV.
- Roossinck MJ, Martin DP, Roumagnac P, 2015. Plant Virus Metagenomics: Advances in Virus Discovery. *Phytopathology* **105**, 716-27.
- Sasaya T, Ishikawa K, Koganezawa H, 2002. The nucleotide sequence of RNA1 of Lettuce big-vein virus, genus Varicosavirus, reveals its relation to nonsegmented negative-strand RNA viruses. *Virology* **297**, 289-97.
- Scheets K, Jordan R, White KA, Hernandez C, 2015. Pelarspovirus, a proposed new genus in the family Tombusviridae. *Archives of Virology* **160**, 2385-93.
- Scholthof KB, Hillman BI, Modrell B, Heaton LA, Jackson AO, 1994. Characterization and detection of sc4: a sixth gene encoded by sonchus yellow net virus. *Virology* **204**, 279-88.
- Scholz MB, Lo CC, Chain PS, 2012. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Current Opinion in Biotechnology* **23**, 9-15.
- Seo JK, Kwak HR, Kim MK, Choi HS, 2015. First Report of Carnation Italian ringspot virus in *Chrysanthemum zawadskii* var. *latilobum* in Korea. *Plant Disease* **99**, 1.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I, 2009. ABySS: a parallel assembler for short read sequence data. *Genome Research* **19**, 1117-23.
- Stobbe AH, Roossinck MJ, 2014. Plant virus metagenomics: what we know and why we need to know more. *Frontiers in Plant Science* **5**, 150.
- Stuart G, Moffett K, Bozarth RF, 2004. A whole genome perspective on the phylogeny of the plant virus family Tombusviridae. *Archives of Virology* **149**, 1595-610.
- Suzuki S, Ono N, Furusawa C, Ying BW, Yomo T, 2011. Comparison of sequence reads obtained from three Next-Generation Sequencing platforms. *PLoS One* **6**, e19534.
- Taliansky ME, Robinson DJ, 2003. Molecular biology of umbraviruses: phantom warriors. *Journal of General Virology* **84**, 1951-60.
- Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S, 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725-9.
- Von Bubnoff A, 2008. *Next-Generation Sequencing: the race is on*. *Cell* **132**, 721-3.
- Wang Q, Ma X, Qian S, *et al.*, 2015. Rescue of a Plant Negative-Strand RNA Virus from Cloned cDNA: Insights into Enveloped Plant Virus Movement and Morphogenesis. *PLoS Pathogens* **11**, e1005223.
- Weier HG, 2013. Bioinformatics for High Throughput Sequencing. *Data Mining in Genomics & Proteomics* **4**, 3.

Zerbino DR, Birney E, 2008. *Velvet*: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**, 821-9.

Zhai Y, Miglino R, Sorrentino R, Masenga V, Alioto D, Pappu HR, 2014. Complete genomic characterization of Eggplant mottled dwarf virus from *Agapanthus* sp. by deep sequencing and de novo assembly. *Journal of Plant Pathology* **96**, 2.