



Universidade Federal do Pampa

UNIVERSIDADE FEDERAL DO PAMPA

CAMPUS SÃO GABRIEL

**PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU* EM CIÊNCIAS
BIOLÓGICAS**

RAÍSSA OCHÔA GOLIN

**ESTUDO DA EVOLUÇÃO DA PROTEASE 3C DE PICORNAVIRIDAE E VÍRUS
PICORNA-LIKE ATRAVÉS DA SUA SEQUÊNCIA PROTEICA E DOMÍNIOS
CONSERVADOS**

DISSERTAÇÃO DE MESTRADO

SÃO GABRIEL, RIO GRANDE DO SUL, BRASIL

2014

RAÍSSA OCHÔA GOLIN

**ESTUDO DA EVOLUÇÃO DA PROTEASE 3C DE PICORNAVIRIDAE E VÍRUS
PICORNA-LIKE ATRAVÉS DA SUA SEQUÊNCIA PROTEICA E DOMÍNIOS
CONSERVADOS**

Dissertação apresentada ao Programa de Pós-Graduação *Stricto Sensu* em Ciências Biológicas da Universidade Federal do Pampa como requisito parcial para obtenção do Título de Mestre em Ciências Biológicas.

Orientador: Prof. Dr. Andrés Delgado Cañedo

São Gabriel

2014

GOLIN, Raíssa Ochôa

Estudo da evolução da protease 3C de Picornaviridae e vírus
Picorna-like através da sua sequência proteica e domínios conservados./

Raíssa Ochôa Golin.

50 folhas;

Dissertação (Mestrado) Universidade Federal do Pampa, 2014.

Orientação: Prof. Dr. Andrés Delgado Cañedo.

1.Genética. 2.Abelhas. 3. Virologia.

I. Delgado-Cañedo, Andrés.

II. Doutor

RAÍSSA OCHÔA GOLIN

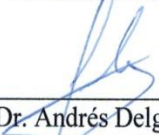
**ESTUDO DA EVOLUÇÃO DA PROTEASE 3C DE PICORNAVIRIDAE E VÍRUS PICORNA-
LIKE ATRAVÉS DA SUA SEQUÊNCIA PROTEICA E DOMÍNIOS CONSERVADOS**

Dissertação apresentada ao Programa de Pós-graduação
Stricto Sensu em Ciências Biológicas da Universidade
Federal do Pampa, como requisito parcial para obtenção
do Título de Mestre em Ciências Biológicas.

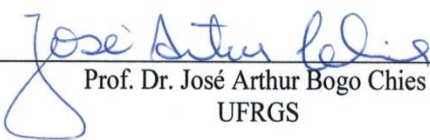
Área de concentração: Genética

Dissertação defendida e aprovada em: 21 de março de 2014.

Banca examinadora:



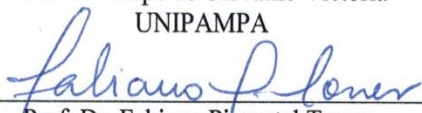
Prof. Dr. Andrés Delgado Cañedo
Orientador
UNIPAMPA



Prof. Dr. José Arthur Bogo Chies
UFRGS



Prof. Dr. Filipe de Carvalho Victoria
UNIPAMPA



Prof. Dr. Fabiano Pimentel Torres
UNIPAMPA

Dedico essa dissertação aos meus pais,
Rosa Nara Ochôa e Paulo Ricardo Golin.

AGRADECIMENTOS

Agradeço, primeiramente, ao meu orientador Andrés Delgado Cañedo, pela oportunidade dada de trabalhar com as abelhas e com a virologia. Hoje posso dizer que descobri o que gosto de pesquisar e isso graças a você! Além de toda paciência e amizade, quero algum dia conseguir retribuir tudo o que tu fizeste por mim! La garantía es usted!

Agradeço a minha família e ao Matheus pelo apoio de sempre! Por confiarem em mim e não deixarem que em algum momento eu desistisse, ou ainda, por aguentarem as crises!

Aos colegas do PPG - Ciências Biológicas! Claudinha, Douglas, Dione, Pábulo, Ana, Rafa, Fernanda e Rayssa! Por "estarmos no mesmo barco" e nos apoiarmos sempre! Aos professores também pelos ensinamentos em sala de aula e em vida!

As Apigirls e ao Maikel! Nome dado pelo nosso chefe! Pelas reuniões de trabalho que acabaram em amizade! Carol, Mari, Vivi, Maikel ...Obrigado! E em especial pra Cissa Barcelos, pelas tardes e noites infundáveis buscando e desbastando sequências, e também pela grande amiga que és!

Ao Leandro Nascimento, vulgo Baraka, e a Lilian Machado pelas ajudas em Bioinformática!

À Escola XV de Novembro, meu local de trabalho, pela cooperação e flexibilidade para que eu pudesse atender as exigências do mestrado.

À todos que de alguma forma colaboraram para a execução dessa pesquisa, esse trabalho é a soma de todos vocês! Se esqueci de citar alguém aqui é por falta de memória e não por falta de amigos! Obrigado!

"Nós vamos morrer, e isso nos torna afortunados. A maioria das pessoas nunca vai morrer, porque nunca vai nascer. As pessoas potenciais que poderiam estar no meu lugar, mas que jamais verão a luz do dia, são mais numerosas que os grãos de areia da Arábia. Certamente esses fantasmas não nascidos incluem poetas maiores que Keats, cientistas maiores que Newton. Sabemos disso porque o conjunto das pessoas possíveis permitidas pelo nosso DNA excede em muito o conjunto de pessoas reais. Apesar dessas probabilidades assombrosas, somos eu e você, com toda a nossa banalidade, que aqui estamos..."

Richard Dawkins

RESUMO

As abelhas apresentam uma combinação de características individuais e ainda de cooperação animal não encontrada no restante do reino animal. São insetos sociais e participam da polinização de diversas plantas que fornecem alimento para o homem. No Brasil com a africanização das abelhas, essas tornaram-se altamente produtoras e enxameadoras, o que vem tornando o país uma potência na produção de mel e outros produtos originados da atividade apícola. Muitas doenças podem afetar as abelhas, dentre elas muitas causadas por vírus. Controlar as infecções virais é essencial para a manutenção ecológica das abelhas e da produção apícola. Ao mesmo tempo, existe a possibilidade de relacionar vírus que infectam humanos com os vírus que infectam as abelhas, visto que os tratamentos utilizados para os seres humanos poderiam ser utilizados em colméias e, ao mesmo tempo, utilizar as abelhas como modelo de estudo para o desenvolvimento de novos antivirais. Na busca por um ponto em comum analisamos filogeneticamente a protease 3C, que ocorre nos vírus da super-família Picorna-like, onde encontram-se os vírus que parasitam as abelhas. Essa protease tem a capacidade de clivar a poliproteína viral nas proteínas maduras do vírus e ainda causar a degradação proteolítica das proteínas do hospedeiro. Até hoje não foi encontrada uma forma de utilizar a protease 3C em estudos filogenéticos pois existe muita divergência das suas sequências entre os vírus. O objetivo dessa pesquisa foi identificar uma forma de analisar filogeneticamente a protease 3C. As sequências da protease 3C e da RdRp de 55 vírus foram coletadas do NCBI (National Center of Biotechnology Information) e submetidas ao MEME (Multiple Em for Motif Elicitation), onde foram obtidos quatro sítios conservados. Após foi realizada a análise filogenética dos sítios conservados por Máxima Verossimilhança e análise da distância entre os sítios por parcimônia e ainda foi construída uma árvore com base na RdRp. As árvores dos sítios 1 e 2 apresentaram uma melhor robustez estatística e agrupamento dos vírus. Essas regiões conservadas da protease 3C-Pro podem ser o início para estabelecermos uma relação entre as proteases dos picornavírus e vírus *picorna-like* na busca da compreensão do seu mecanismo de infecção viral e também uma alternativa de estudo para outras sequências com alta variabilidade. O uso dos domínios 1 e 2 proporcionou a árvore com maior robustez apresentada até o dia de hoje para esta proteína viral.

Palavras-chave: Protease 3C - Vírus - Abelhas - Picorna-like - regiões conservadas.

ABSTRACT

The bees have a combination of individual features and animal cooperation is not yet found in the rest of the animal kingdom . They are social insects and participate in the pollination of many plants that provide food for man . In Brazil with the africanization of bees , these have become highly producing and swarm , which is making the country a power in the production of honey and other products derived from beekeeping . Many diseases can affect bees , among them many caused by viruses . Control viral infections is essential for ecological maintenance of bees and beekeeping . At the same time , it is possible to relate viruses that infect humans and viruses that infect the bees , whereas the treatments for humans could be used in beehives and at the same time using the bees as a model for development of new antiviral agents. In the search for a common point analyzed phylogenetically 3C protease , which occurs in the superfamily Picorna -like, which are viruses that parasitize bees virus. This protease is capable of cleaving the polyprotein, the mature viral proteins and viruses also cause the proteolytic degradation of host proteins . Until today there a way to use the 3C protease was found in phylogenetic studies because there is much divergence of their sequences between virus. The objective of this research was to identify a way to analyze phylogenetically 3C protease . The 3C protease and RdRp sequences of 55 viruses were collected from the National Center for Biotechnology Information , submitted to MEME , where four conserved sites were obtained . Upon phylogenetic analysis of the conserved sites and by maximum likelihood analysis of the distance between sites by parsimony was performed and was still a tree constructed based on the RdRp . Trees of sites 1 and 2 had a better statistical robustness and clustering of virus. These conserved regions of the 3C protease - Pro may be the start to establish a relationship between the proteases of the picornavirus and Picorna-like viruses in the quest to understand the mechanism of viral infection and also an alternative study for other sequences with high variability . The use of domains 1 and 2 provided the tree with greater robustness displayed until today for this viral protein.

Key-words: Protease 3C - viruses - bees - Picorna-like - conserved regions.

LISTA DE FIGURAS

| | |
|---|-----------|
| Figura 1. Estrutura do cápside da família Picornaviridae com os protômeros..... | 3 |
| Figura 2. Organização do genoma dos picornavírus..... | 4 |
| Figura 3. Mapa do genoma do DWV mostrando os sítios consenso de clivagem pela 3C - Pro..... | 5 |
| Figura 4. Organização do genoma do complexo KIA com indicação do sítios consenso de clivagem pela 3C -Pro..... | 5 |
| MANUSCRITO | |
| Figura 1. Fluxograma dos procedimentos utilizados na coleta dos dados | 15 |
| Figura 2. Aminoácidos conservados nas áreas selecionadas pelo MEME | 16 |
| Figura 3. Árvore das distâncias entre os sítios conservados analisada por parcimônia | 24 |
| Figura 4. Árvore dos motivos 1 e 2 construída por Máxima Verossimilhança | 25 |
| Figura 5. Árvore da RdRp construída por Máxima verossimilhança com os padrões de distâncias | 26 |
| Figura 6. Árvore do motivo 1 que não foi incluída no manuscrito | 35 |
| Figura 7. Árvore do motivo 3 que não foi incluída no manuscrito | 36 |
| Figura 8. Árvore do motivo 4 que não foi incluída no manuscrito | 37 |

LISTA DE ABREVIATURAS

- ABPV- (Acute bee paralysis virus) - Vírus da paralisia aguda das abelhas
- BQCV- (Black queen cell virus) - Vírus da realeira negra
- CBPV- (Chronic bee paralysis virus) - Vírus da paralisia crônica das abelhas
- CCD- (Colony collapse disorder) - Doença do colapso de colônias
- CWV- (Cloud wing virus) -
- DNA- (Desoxirribonucleic acid) - Ácido desoxirribonucleico
- DWV- (Deformed wing virus) - Vírus deformador da asa
- FV- (Filamentous virus) - Vírus filamentosos
- IAPV- (Israeli acute paralysis virus) - Vírus da paralisia aguda israelita
- KBV- (Kashmir bee virus) - Vírus caxemira das abelhas
- KIA- Complexo formado pelos dicistrovirus Kashmir bee virus, Israeli acute paralysis virus e Acute bee paralysis virus
- KV- (Kakugo virus) - Vírus kakugo
- ML- (Maximum likelihood) - Máxima verossimilhança
- m-RNA- RNA mensageiro
- ORF- (Open reading frame) - sequência de leitura aberta
- RdRp - RNA dependente de RNA polimerase
- RNA- (Ribonucleic acid) - Ácido ribonucleico
- SBV- (Sacbrood bee virus) - Vírus da cria ensacada das abelhas
- SBPV- (Slow bee paralysis virus) - Vírus da paralisia lenta das abelhas
- VDV- (Varroa destructor virus) - Vírus da varroa destructor
- +ssRNA- RNA simples fita com polaridade positiva
- 3C-Pro - protease 3C

APRESENTAÇÃO

No item INTRODUÇÃO, consta uma breve revisão da literatura sobre o tema abordado nesta dissertação.

No item MANUSCRITO, constam os materiais e métodos e respectivos resultados e discussão da dissertação na forma de artigo que será submetido para o Journal of Virology.

O item CONSIDERAÇÕES FINAIS, encontrado no final desta dissertação, apresenta interpretações e comentários gerais sobre os resultados presentes neste trabalho.

As REFERÊNCIAS referem-se às citações utilizadas no item INTRODUÇÃO dessa dissertação.

Nos ANEXOS dessa dissertação encontram-se as árvores filogenéticas que não foram colocadas no artigo.

SUMÁRIO

| | |
|---|-------------|
| RESUMO..... | iv |
| ABSTRACT | v |
| LISTA DE FIGURAS..... | vi |
| LISTA DE ABREVIATURAS..... | vii |
| APRESENTAÇÃO | viii |
| SUMÁRIO | ix |
| 1 INTRODUÇÃO | 1 |
| 1.1 As abelhas | 1 |
| 1.2 Vírus que parasitam as abelhas | 2 |
| 1.3 Família Picornaviridae e vírus Picorna-like..... | 2 |
| 1.4 Por que estudar a Protease 3C? | 4 |
| 1.5 Colony Collapse Disorder | 6 |
| 1.6 Análise Filogenética da Superfamília Picorna-like..... | 6 |
| 1.7 Justificativa..... | 8 |
| 1.8 Problema..... | 8 |
| 2. OBJETIVO DA PESQUISA..... | 9 |
| 3. MANUSCRITO | 10 |
| 4 CONSIDERAÇÕES FINAIS..... | 30 |
| 5 REFERÊNCIAS BIBLIOGRÁFICAS | 31 |
| ANEXOS | 34 |

1 INTRODUÇÃO

1.1 As abelhas

É incontestável a importância das abelhas para a natureza. As abelhas apresentam uma combinação de características individuais e ainda de cooperação animal não encontrada no restante do reino animal. Além de nos fornecer mel, cera, própolis, geléia real e pólen, elas polinizam várias plantas que suprem a humanidade de alimentos. Segundo Wiese (2005) as abelhas são responsáveis por cerca de 80 a 100% da produção agrícola de sementes e frutas. Além disso, desempenham papel crucial na manutenção da biodiversidade, uma vez que polinizam várias plantas que requerem obrigatoriamente a polinização cruzada para sua reprodução (Allen-Wardell et al., 1998).

No Brasil, a primeira abelha do gênero *Apis* a ser introduzida foi a *Apis mellifera mellifera* L., conhecida como abelha alemã. Foi importada pelo Padre Antônio Carneiro em 1839 para o Rio de Janeiro. Frederico A. Hanemann trouxe para o Rio Grande do Sul, entre 1870 e 1880, a abelha italiana *Apis mellifera ligustica* L.. Provavelmente as espécies *Apis mellifera carnica* L. e *Apis mellifera caucasica* L. também foram introduzidas na mesma época. A abelha africana *Apis mellifera scutelata* L. só foi introduzida em 1956, em Rio Claro, São Paulo, pelo Dr. Warwick E. Kerr, que começou a fazer estudos com essa espécie. Em decorrência de um acidente, enxames conseguiram fugir e então houve o cruzamento da abelha africana com as espécies já existentes no país, consideradas mansas. Desta forma originaram-se as abelhas africanizadas, que não são mansas e possuem grande enxameação. Esta última característica facilitou que estas se espalhassem por todo o território nacional e todo o continente americano, com a exceção de regiões muito frias.

Uma colméia pode chegar a ter 60.000 operárias, 400 zangões e 1 rainha. As abelhas africanizadas se reproduzem com mais intensidade, como também abandonam facilmente a colméia caso houver falta de alimento, muito frio, umidade, doenças, etc. Porém, independentemente da causa ser a seca, o frio ou outro fator ambiental, quando as abelhas se encontram em condições desfavoráveis começam a sofrer as consequências do estresse, o que se traduz no enfraquecimento do sistema imunológico o que conseqüentemente as predispõem à infecção por uma série de patógenos (Oldroyd, 2007). Como as colméias são extremamente populosas, todo e qualquer tipo de parasita que infectar uma única abelha poderá comprometer a colméia inteira, em função do grande contato existente entre abelhas. Esses

parasitas podem ainda parasitar todas as castas (operárias, zangões e rainha) como também todos os estágios de vida (ovo, larva, pupa e adulto).

1.2 Vírus que parasitam as abelhas

A infecção viral é o maior fardo para todos os organismos eucarióticos. Como consequência, estes possuem mecanismos eficientes para combater infecções virais que diferem em vertebrados, insetos e plantas (Azzami *et al.*, 2011). No que tange aos vírus conhecidos que infectam as abelhas estes são de 18 tipos diferentes e podem pertencer a 4 Famílias: Cripaviridae, Dicistroviridae, Iflaviridae e Iridoviridae. As três primeiras famílias possuem RNA no seu genoma e a última DNA (Genersch, 2010). Até o final dos anos 1980, as infecções virais específicas das abelhas melíferas eram geralmente consideradas inofensivas em todos os países. Hoje os vírus são considerados um dos fatores responsáveis pela diminuição da produção de mel e desaparecimento das abelhas. Uma colméia pode estar infectada por vírus sem demonstrar sintomas aparentes, o que dificulta o tratamento. Como esses insetos vivem em sociedade, isso de certa maneira, facilita a infecção, sobrevivência e disseminação de parasitas (Fries e Camazine, 2001).

Muitas doenças causam impacto negativo sobre populações de abelhas selvagens e domesticadas, causando graves perdas econômicas. Dos 18 vírus que foram identificados em abelhas (Allen & Ball, 1996), os mais estudados são: Black Queen Cell Virus (BQCV), Deformed Wing Virus (DWV), Filamentous Virus (FV), Chronic Bee Paralysis Virus (CBPV), Acute Bee Paralysis Virus (ABPV), Israeli Acute Paralysis Virus (IAPV), Cloud Wing Virus (CWV), Kashmir Bee Virus (KBV), Sacbrood Virus (SBV), Kakugo Virus (KV), Varroa Destructor Virus (VDV). Segundo Shen *et al.* (2005) a transmissão de doenças na sociedade das abelhas pode servir como um sistema modelo para a transmissão de doenças na sociedade humana e este fato torna importante o estudo destes vírus porque poderíamos estar contribuindo tanto na saúde da abelha e produtividade das suas colmeias como também no conhecimento de mecanismos que permitam diminuir ou controlar infecções virais em humanos.

1.3 Família Picornaviridae e vírus Picorna-like

A família Picornaviridae contém cinco gêneros: Enterovirus, Rhinovirus, Cardiovirus, Aphotavirus, Hepatovirus (Lin *et al.*, 2009). Os vírus desta família, comumente chamados de

picornavírus, são vírus que possuem como material genético um RNA simples fita com polaridade positiva (+ssRNA) e infectam animais domésticos, silvestres, plantas e o ser humano. A informação genética contida na sua simples fita de RNA com sentido positivo é expressa em uma única proteína de, aproximadamente, 2000 aminoácidos (Seipelt *et al.*, 1999). Esta poliproteína gerada é posteriormente clivada em proteínas virais maduras pelas suas proteases, contidas nas respectivas poliproteínas. Segundo Nakashima e Nakamura (2008), a caracterização dos produtos dos sítios de clivagem da ORF1 facilitaria a análise de como a tradução e síntese de proteína do hospedeiro é controlada pela infecção viral.

Os vírions consistem de um cápside com simetria icosaédrica, não-envelopado, de 22-30 nm de diâmetro, sem projeções. O cápside é composto de 60 unidades idênticas, chamadas de protômeros, cada um formado por três proteínas na superfície externa (VP1, VP2 E VP3) e, na maioria dos picornavírus, uma proteína interna (VP4), conforme é ilustrado na figura 1.

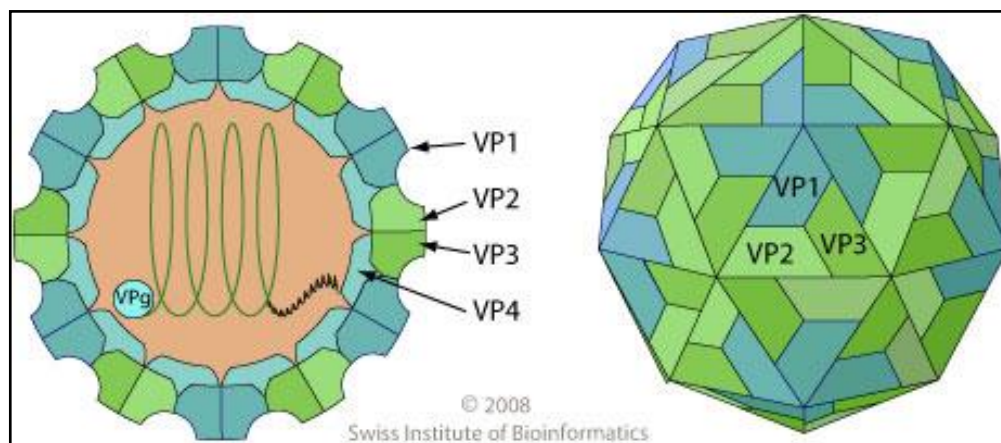


Fig. 1 Estrutura do cápside da família Picornaviridae com os protômeros.

Fonte: <http://viralzone.expasy.org>

A sua molécula de RNA possui uma única sequência de leitura aberta, a extremidade 3' é poli adenilada e a extremidade 5' não possui CAP; porém, é protegida por uma proteína pequena, denominada VPg, que se liga de forma covalente à extremidade, conforme ilustra a figura 2. O RNA viral é infeccioso por si só e têm função de RNA genômico e de RNA mensageiro (m-RNA).

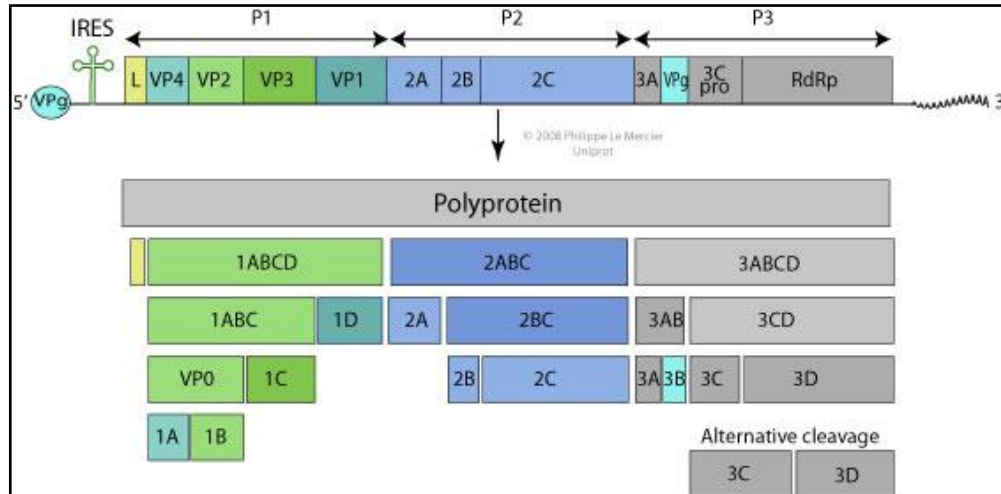


Figura 2. Organização do genoma dos picornavírus.

Fonte: <http://viralzone.expasy.org>

Os vírus das famílias Dicistroviridae e Iflaviridae são denominados Picorna-like, pois tem funcionamento semelhante aos vírus da família Picornaviridae. Os vírus de RNA parasitas específicos das abelhas encontram-se nessas duas famílias. Deformed wing virus (DWV), Varroa destructor virus (VDV), Kakugo virus (KV), Sacbrood virus (SBV), Slow bee paralysis virus (SBPV) são classificados como iflavírus e, Israeli acute paralysis virus (IAPV), Kashmir bee virus (KBV), Acute bee paralysis virus (ABPV) e Black queen cell virus (BQCV) são classificados como dicistrovírus (Baker and Schroeder, 2008).

1.4 Por que estudar a Protease 3C?

Proteases (proteínases, peptidases ou enzimas proteolíticas) são enzimas que quebram ligações peptídicas entre os aminoácidos das proteínas. O processo é chamado de clivagem proteolítica, um mecanismo comum de ativação ou inativação de enzimas. A proteína 3CD é a precursora da protease 3C madura e da polimerase 3D (Harris *et al.*, 1992).

A protease 3C foi encontrada pela primeira vez no vírus da encefalomiocardite (EMCV), e posteriormente mostrou-se ser codificada por todos os picornavírus. Foi designada como proteinase 3C (3C Pro) e é responsável pela realização da maioria dos eventos de clivagem na poliproteína picornaviral. A caracterização destas proteinases é realizada pelo seu mecanismo catalítico de especificidade de substrato e pela sua estrutura tridimensional (Seipelt *et al.*, 1999). Contudo, além de clivar a poliproteína viral, essa protease participa da regulação das funções das células hospedeiras através da degradação proteolítica de proteínas

do hospedeiro moldando o ambiente celular para facilitar a replicação viral e evitar os mecanismos antivirais (Wang e Chen, 2007).

No caso de vírus que infectam abelhas, os sítios de clivagem da poliproteína do vírus DWV foram identificados (Lanzi *et al.* 2006) conforme ilustra a Figura 3.

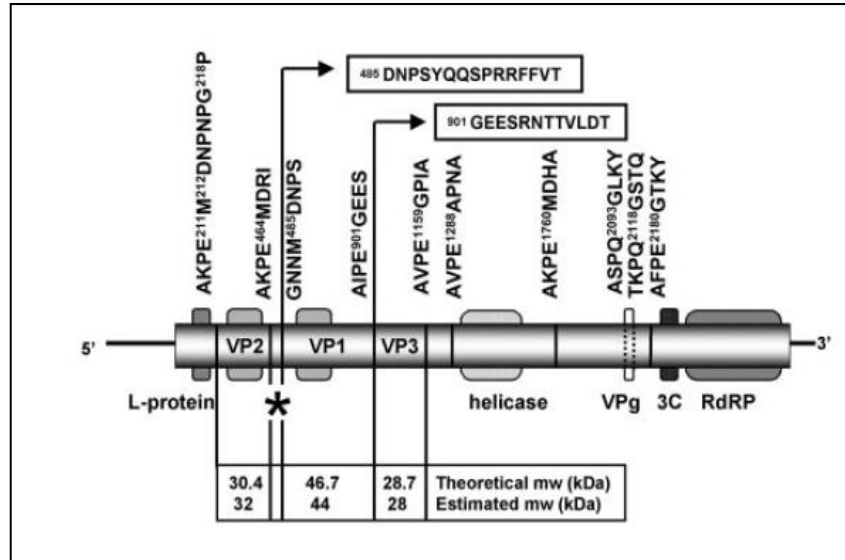


Figura 3. Mapa do genoma do DWV mostrando os sítios de clivagem consenso pela 3C - Pro (Lanzi *et al.* 2006).

Um estudo semelhante foi realizado por Miranda *et al.*, (2010) para os vírus KBV, IAPV e ABPV denominado complexo KIA, conforme ilustra a figura 4.

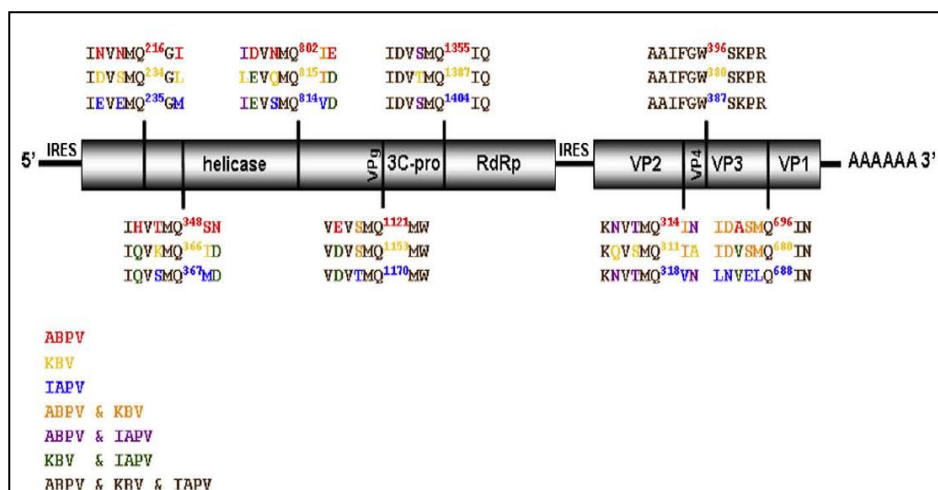


Figura 4. Organização do genoma do complexo KIA com indicação do sítios consenso de clivagem pela 3C -Pro(Miranda *et al.* 2010).

Através do conhecimento desses sítios de clivagem específicos para a protease 3C-Pro, podemos identificar potenciais proteínas alvo e estudar as estratégias que os vírus utilizam para infectar seu hospedeiro, nesse caso as abelhas.

1.5 Colony Collapse Disorder

Os vírus são uma das principais causas da morte das abelhas e estão intimamente relacionados com o fenômeno conhecido como Doença do Colapso das Colônias, Colony Collapse Disorder (CCD), que vem causando perdas milionárias para os apicultores de todo o mundo. O fenômeno do CCD foi relatado pela primeira vez em 2006, no entanto, os apicultores notaram declínios consistentes nas colônias já em 2004 (Cox-Foster, D. L. et al., 2007). CCD é caracterizado pela rápida perda da população de abelhas adultas de uma colônia sem motivos aparentes (Oldroyd, 2007; Stokstad, 2007).

Com base em evidências de que o CCD é transmissível através da reutilização de equipamentos, Pettis et. al (2007), trabalharam sob a hipótese de que o CCD é devido à introdução de um agente infeccioso não reconhecido anteriormente. Várias pesquisas estão sendo realizadas ao redor do mundo buscando explicar o CCD, apontando para efeitos isolados ou sinérgico de patógenos. Nesse trabalho buscou-se considerar o CCD como um efeito sinérgico dos vírus parasitas de abelhas e da sua forma de infecção, a partir das potenciais proteínas por eles clivadas.

1.6 Análise Filogenética da Superfamília Picorna-like

Atualmente a filogenética não busca somente elucidar relações hierárquicas, visando também investigar relações de espécies que interagem entre si, tais como hospedeiros e parasitas ou simbioses, co-evolução de insetos e plantas, etc. Segundo Schneider (2007), esta expansão no uso das ferramentas da filogenética é sem dúvida decorrente da coevolução entre microcomputadores e programas de análise filogenética, cada vez mais complexos.

Existem vários métodos de reconstrução de árvores filogenéticas, dentre eles o método de Máxima Verossimilhança (Maximum Likelihood ML). Os valores de verossimilhança são convertidos em logaritmos para toda árvore e a árvore que apresentar o maior escore de verossimilhança será escolhida. A suposição é que a árvore com maior probabilidade de atingir o estado observado é preferida à árvore com mais baixa probabilidade. Exige um

grande esforço computacional, o que faz com que métodos não exaustivos sejam utilizados também.

Um método utilizado como simplificação do método de Máxima Verossimilhança é o método de Parcimônia, o qual foi empregado por Eck e Dayhoff (1966) para a análise de sequências moleculares. Esse método considera que se existe mais de uma explicação para uma dada observação, devemos adotar aquela mais simples, o que acaba por minimizar os passos evolucionários para explicar o conjunto de dados. Porém, o número de árvores possíveis pode atingir valores muito altos, devendo ser criada uma árvore consenso.

Tanto a Máxima verossimilhança quanto a Parcimônia são métodos que podem ser utilizados para a análise e comparação de nucleotídeos como também de aminoácidos. Antes da invenção de métodos rápidos de sequenciamento de DNA em 1977(Maxam and Gilbert 1977; Sanger et al. 1977), muitos estudos de evolução molecular foram conduzidos usando banco de dados de sequencia de aminoácidos. Segundo Nei e Kumar (2000), sequências de aminoácidos ainda são úteis para estudos evolucionários, sendo mais conservadas que as sequências de DNA, tornando-se útil para o estudo da evolução dos genes ou das espécies.

A superfamília dos vírus picorna-like é em particular caracterizada pela conservação parcial de um conjunto de genes que codificam a RNA polimerase dependente de RNA (RdRp), protease quimiotripsina - like (3C-Pro), a superfamília da helicase 3 (S3H) e uma proteína de ligação ao genoma (chamada proteína viral - VPg) (Koonin, 2008). Porém, somente as regiões que codificam as proteínas RdRp e S3H são utilizadas na maioria dos estudos de classificação e evolução dos picornavirus a partir dos estudos de Koonin e Dolja (1993). Até hoje não se tem uma relação filogenética baseada na protease 3C-Pro, pois segundo Gorbalenya e colaboradores (1989), devido à falta de semelhança na sequência total da 3C-Pro com as 3C-like proteases, foi sugerido que elas não possuem relação evolutiva com as proteases de cisteína, sendo a sua analogia explicada por convergência.

Embora não seja muito conservada evolutivamente a 3C-pro é muito importante para o processo infeccioso uma vez que os picornavírus codificam uma poliproteína que posteriormente será clivada pela protease nas suas proteínas funcionais. A análise através dos métodos de Máxima Verossimilhança e Parcimônia podem fornecer dados importantes para a compreensão da atividades dessa protease, tanto na clivagem da sua poliproteína quanto na clivagem das proteínas do hospedeiro.

1.7 Justificativa

O Brasil vem se destacando na produção de mel, tendo como um dos maiores colaboradores o Rio Grande do Sul. Desta forma, é de extrema importância o monitoramento da sanidade dos apiários da região para a incidência de patógenos. No que se refere aos vírus, identificar os seus mecanismos de infecção e ação torna-se necessário para a melhoria da sanidade apícola e conseqüentemente da produção melífera.

Os vírus que parasitam as abelhas são pertencentes a super-família Picorna-like. Desta forma, analisar a protease 3C filogeneticamente pode colaborar para o controle e identificação de infecções causadas por esses vírus e fornecer subsídios para o desenvolvimento de antivirais.

1.8 Problema

O problema que a pesquisa busca responder é: qual a relação filogenética entre os vírus da família Picornaviridae e Picorna-like com base na 3C protease, sendo essa um importante fator para o mecanismo de infecção viral.

2. OBJETIVO DA PESQUISA

O objetivo dessa pesquisa foi identificar uma forma de analisar filogeneticamente a protease 3C, contribuindo para a solução da relação filogenética entre os vírus da família Picornaviridae e vírus Picorna-like, dentre estes os vírus que parasitam as abelhas (*Apis mellifera* L.).

3. MANUSCRITO

PHYLOGENETIC AND STRUCTURAL ANALYSIS OF THE PROTEASE 3C OF THE PICORNA-LIKE SUPERFAMILY THROUGH YOUR 4 MOTIFS CONSERVED*

*Manuscript that will be submitted to Journal of Virology (ISSN 1098-5514)

ABSTRACT

The viruses of the Picornaviridae family and Picorna-like virus represent the largest class of known viruses, having a broad host range. The strategy of the picornavirus infection is based on protease 3C-Pro, which is capable of cleavage the viral polyprotein, as well as host proteins. Most picornavirus phylogenetic studies are performed based on the RdRp and S3H protein sequence and there is a gap for the 3C-Pro based in phylogeny. By the time, phylogenetic studies of this protease brought no conclusive data due to the large variability in 3C-pro protein sequences. In this study we have determined conserved domains, observing 4 motifs and also, through distance between the motif, was established standards for viral 3C-pro proteases. By determining conserved sites Phylogenetic trees were constructed by the method of maximum likelihood and even a Parsimony tree using the distances between the conserved motif. The region of conserved motifs 2 and 1 of the phylogenetic tree demonstrated a better robustness, with the best clustering of families and having a greater resemblance to the tree constructed from the RdRp. On other hand, we submit 3CPro sequences at Conserved Domain Database (CDD). The sequences were recognizes as pfam00548 (present in 25 of the studied viruses) wens motif 1 and 2 were separated by 3 aminoacids, when separation was larger this domain was not detected, in these case were obtained the domains with cl02893 code (present in 16 of the studied viruses), cl13774 (present in 7 of the studied viruses) or it was not possible to detected any type of peptidase domain (7 the studied viruses). The tree obtained using inter-motif distances shown a classification where most viruses grouped in with your family, but mixing family members of Iflaviridae and Dicistroviridae and some picornavirus. These conserved regions of the protease 3C-Pro may be the start to establish a relationship between proteases of the picornavirus and picorna-like viruses to understand the mechanism of viral infection and also an alternative study for other sequences with high variability. We based on our data propose the use of motif 1 and 2 weigh matrix provided a great tool for recognize 3C-pro in the Picorna-like virus superfamily and, also, the motif 1 and 2 sequences and inter-motif distances in 3C-pro phylogenetic analysis .

Key - words: Evolution - Picornaviridae, 3C - Pro - Sites Conserved - Phylogenetics.

INTRODUCTION

Viruses from Picornaviridae family are small, with genome consisting of single-stranded RNA positive sense (between 7000 and 8500 bases) (2). The viral genome is

translated into a single polyprotein which is processed by the viral protease autocleavage generating the functional proteins (1). They are classified in Class IV (7), which is the largest class of existing viruses (8), also, in this class are the Dicistroviridae and Iflaviridae families, as part of the super-family Picorna-like (Order Picornavirales). Despite their similarities, there are some differences when the genomic organization is compared. Picornaviridae and Iflaviridae are composed of a single open reading frame (ORF) encoding a single polyprotein containing the information for structural and non-structural protein sequence (2,17,14). In Dicistroviridae, its genome consists of two ORFs separated by an internal ribosomal entry site (IRES) sequence. ORF1 encodes non-structural proteins helicase, 3C protease and RNA-dependent RNA polymerase, whereas the ORF2 encodes four capsid proteins (10). These families include viruses that cause from simple colds to hepatitis A and polio and can infect domestic and wild animals, insects, plants and man.

The superfamily of Picorna-like virus is particularly characterized by the partial retention of a set of genes encoding the RNA dependent RNA polymerase (RdRp), protease chymotrypsin-like (3C-Pro), the helicase 3 superfamily (S3H) and a protein binding to the genome (called viral protein - VPg) (13).

Although not very conserved evolutionarily 3C-pro is very important for the infectious process since picornaviruses encode polyprotein must be subsequently cleaved by the protease in its functional proteins, where each 3C-Pro is specific for its own polyprotein (11). These viral proteases have a catalytic cysteine that replaces the active serine of other proteases and also a termination of trypsin-like protease (24). Beyond their essential role in the processing and maturation of the viral polyprotein, the viral 3C-Pro may participate in regulating functions of the host cells by proteolytic degradation of host proteins, shaping the cellular environment to facilitate viral replication and prevent the antiviral mechanisms (11,23,24,21,16).

In concern to the evolution, the regions encoding either RdRp or S3H proteins are well conserved and, therefore, are used in the classification and evolution of picornavirus since Dolja and Koonin studies (12). On the other hand, 3C-Pro sequence is very variable and, until now, is more common than S3H and could be found in Picornavirus families and picornavirus-like viruses as well as other Calciviridae, Comoviridae, Dicistroviridae, Iflaviridae, Marnaviridae and Sequiviridae (13).

However, so far, not was possible to find a solution to the phylogenetic relationship of the 3C-Pro protease between picorna-like superfamily. In this work we use the strategy of determining conserved motifs in 3C Pro proteins in members of the Picornaviridae Family and in viruses of the families considered Picorna-like (Iflaviridae and Dicistroviridae). Also we analyzed these motif both from the distances that separate these sequences or constructing phylogenetic tree from their sequences, demonstrating that even the phylogenetic trees constructed from statistically conserved motifs did not shown similar statistical robustness when compare with RdRp protein trees, but it has presented the most robust tree obtained until now for 3C-pro protease. At same time, the phylogenetic tree based on the distances between the conserved domains, demonstrated similarities when compared to the RdRp tree, obtaining a new strategy to determine and study the relationship between 3C pro of Picornavirus and picornavirus-like viruses.

MATERIALS AND METHODS

Sequence Data. The sequences of the protease 3C - Pro were obtained from the National Center for Biotechnology Information (18). 55 viruses were used grouped in the families: Picornaviridae (33 viruses), Iflaviridae (14 viruses), Dicistroviridae (7 Virus) and 1 (Unassigned) use an outgroup (*Rhizosolenia setigera*). Some sequences were obtained directly as the 3C-pro processed form, other as viral polyprotein with annotation for 3C-Pro and when not annotated we obtained the sequence from the identification of conserved motifs using the complete polyprotein as a basis (Table 1).

Table 1. Viruses used in the study with their accession numbers in NCBI. The index of GenBank Accession indicate the way in which the sequence of 3C protease was obtained: ¹ corresponds to the sequence deposited 3C protease; ² corresponds to the sequence deposited polyprotein and has noted that the region of the 3C protease; ³ deposited polyprotein sequence corresponds to 3C region without annotated.

| Virus | Access number | Host | Number of amino acids in the sequence of the protease 3C | Family |
|---------------------------|---------------------------|-------|--|-----------------|
| Acute bee paralysis virus | NP_066241. 1 ² | Bee | 209 | Dicistroviridae |
| Aichivirus A | NP_740443. 1 ¹ | Human | 190 | Picornaviridae |
| Aichivirus B | NP_859027. 1 ¹ | Human | 190 | Picornaviridae |

| | | | | |
|--|-----------------------------|------------------|-----|-----------------|
| Aphid lethal paralysis virus | AGF84786.1 ² | Aphid | 226 | Dicistroviridae |
| Avian encephalomyelitis virus | NP_705604.1 ¹ | Chicken | 211 | Picornaviridae |
| Avian sapelovirus | YP_164830.1 ¹ | Duck | 183 | Picornaviridae |
| Bat picornavirus 1 | YP_004782554.1 ¹ | Bat | 184 | Picornaviridae |
| Bat picornavirus 2 | YP_004782568.1 ¹ | Bat | 184 | Picornaviridae |
| Bat picornavirus 3 | YP_004782540.1 ¹ | Bat | 184 | Picornaviridae |
| Black queen cell virus | ABS82427.1 ² | Bee | 310 | Dicistroviridae |
| Brevicoryne brassicae picorna-like virus | YP_001285409.1 ² | Aphid | 190 | Iflaviridae |
| Canine picornavirus | YP_005352654.1 ¹ | Dog | 184 | Picornaviridae |
| Cosavirus A | ACL15185.1 ² | Human | 189 | Picornaviridae |
| Deformed wing virus | AGA20422.1 ² | Bees | 213 | Iflaviridae |
| Drosophila C virus | NP_044945.1 ² | Fly | 226 | Dicistroviridae |
| Echovirus E6 | AAA65044.1 ² | Human | 166 | Picornaviridae |
| Ectropis obliqua picorna-like virus | AAQ64627.1 ³ | Moth | 187 | Iflaviridae |
| Encephalomyocarditis virus | ACQ90253.1 ² | Swine | 187 | Picornaviridae |
| Enterovirus A | NP_042242.1 ² | Human | 183 | Picornaviridae |
| Enterovirus B | NP_740546.1 ¹ | Human | 183 | Picornaviridae |
| Equinus rhinitis B virus 2 | NP_740570.1 ¹ | Horses | 204 | Picornaviridae |
| Feline picornavirus | YP_004934011.1 ² | Cat | 183 | Picornaviridae |
| Foot-and-mouth disease virus - type O | NP_740466.1 ¹ | Cattle and swine | 207 | Picornaviridae |
| Formica exsecta virus | YP_008888537.1 ³ | Ant | 208 | Iflaviridae |
| Halyomorpha halys virus | YP_008719809.1 ³ | Bedbug | 198 | Iflaviridae |
| Hepatitis A virus | 1 HAV ¹ | Human | 191 | Picornaviridae |
| Human cosavirus E1 | ACL99815.1 ² | Human | 189 | Picornaviridae |
| Human coxsackievirus A24 | ADZ57561.1 ¹ | Human | 183 | Picornaviridae |
| Human coxsackievirus B4 | AFR79234.1 ² | Human | 183 | Picornaviridae |

| | | | | |
|-------------------------------------|-----------------------------|-----------------|-----|-----------------|
| Human enterovirus C109 | YP_003795184.1 ¹ | Human | 183 | Picornaviridae |
| Human parechovirus 1 | ABK54353.1 ² | Human | 183 | Picornaviridae |
| Human parechovirus 6 | ABS82455.1 ² | Human | 183 | Picornaviridae |
| Human poliovirus 1 | ACF16830.1 ² | Human | 183 | Picornaviridae |
| Human poliovirus 2 | AAO88056.1 ² | Human | 183 | Picornaviridae |
| Infectious flacherie virus | NP_620559.1 ³ | Moth | 208 | Iflaviridae |
| Israeli acute paralysis virus | YP_001040002.1 ² | Bees | 219 | Dicistroviridae |
| Kakugo virus | YP_015696.1 ² | Bees | 213 | Iflaviridae |
| Kashmir bee virus | NP_851403.1 ² | Bees | 222 | Dicistroviridae |
| Ljungan virus | NP_705883.1 ¹ | Rodents | 211 | Picornaviridae |
| Lygus lineolaris virus | AEL30247.1 ² | Bedbug | 193 | Iflaviridae |
| Nilaparvata lugens honeydew virus 1 | BAN19725.1 ³ | Aphid | 218 | Iflaviridae |
| Perina nuda virus | NP_277061.1 ³ | Moth | 185 | Iflaviridae |
| Pigeon picornavirus B | YP_004564600.1 ² | Pigeon | 184 | Picornaviridae |
| Rhinovirus A | NP_740400.1 ¹ | Human | 183 | Picornaviridae |
| Rhinovirus C | YP_001552441.1 ¹ | Human | 183 | Picornaviridae |
| Rhizosolenia setigera RNA virus 01 | YP_006732323.1 ² | Bacillariophyta | 179 | Unassigned |
| Sacbrood virus | AFX65180.1 ² | Bees | 212 | Iflaviridae |
| Simian enterovirus 46 | AAL69636.2 ² | Simian | 184 | Picornaviridae |
| Simian sapelovirus 1 | NP_937978.1 ¹ | Simian | 182 | Picornaviridae |
| Slow bee paralysis virus | YP_003622540.1 ² | Bees | 201 | Iflaviridae |
| Spodoptera exigua iflavirus 1 | YP_004935363.1 ³ | Moth | 203 | Iflaviridae |
| Swine pasivirus 1 | YP_006543639.1 ² | Swine | 195 | Picornaviridae |
| Taura syndrome virus | AEQ61915.1 ² | Shrimp | 203 | Dicistroviridae |
| Theiler's encephalomyelitis virus | AAA47930.1 ² | Mice | 202 | Picornaviridae |
| Varroa destructor virus 1 | AEM63700.1 ² | Bees | 210 | Iflaviridae |

Analysis of conserved sites. In order to detect conserved motifs, the sequences were subjected to multiple at Motif is Elucitation 4.9.1 (MEME) (5) using as analysis parameter the occurrence of at least one sequence motive by sequence, minimum width $> = 10$, maximum width = 50 and Maximum number of 10 motifs. The motifs obtained were submitted to MAST tool (4) to locate the areas in the sequences analyzed and subsequent thinning from the beginning of the motif 4 until the final of the motif 1. At the same time, all sequences were subjected to CDD data (Conserved Domain Database) (15) to determine wich protease domain could be recognized .

Phylogenetic analysis. From the conserved domains obtained was performed a matrix with the distances wich separate these areas and with this data we constructed a parsimony tree analysis. Moreover, the sequences of the conserved domains were used individually or in combination to construct phylogenetic trees using the maximum likelihood statistical analysis with 5000 bootstrap replicates in Mega 6.06 (22) software, using as outgroup the Rhizosolenia setigera RNA virus, that parasites bacillariophyta and has not been classified in any of the group IV family of Baltimore classification (7). For tree motif 2 e 1 and tree of RdRp, Mega software chose the JTT evolutionary model with gamma distribution.

Comparison of 3C-Pro and RdRp. To compare the obtained data we constructed a maximum likelihood tree of the RdRp with 1000 bootstrap replicates. The entire procedure may be observed in the flowchart (Figure 1). After identification of conserved sites and construction of the trees, the Aphid lethal virus was excluded from parsimony analysis by not presenting the motif 3. For parsimony analysis of the Mesquite software was used.

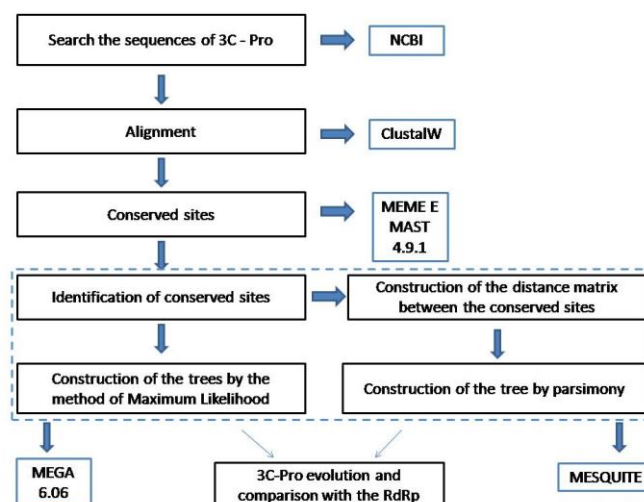


Figure 1. Flowchart of procedures used in data collection.

RESULTS

Determination of conserved motifs. Koonin *et al.* (13), in previous studies have tried to establish a phylogenetic relationship of 3C protease using its complete sequence, which brought no significant data due to low information content in the multiple sequence alignment. Thus, we seek a solution to this problem by working only with the conserved regions of the sequences of the 3C protease. After submitting sequences to the MEME tool, we obtained four motifs statistically conserved and shared by the 55 viruses studied (Figure 2). The weight matrix of the fields used by the program to make the search can be seen in the supplementary material 1.

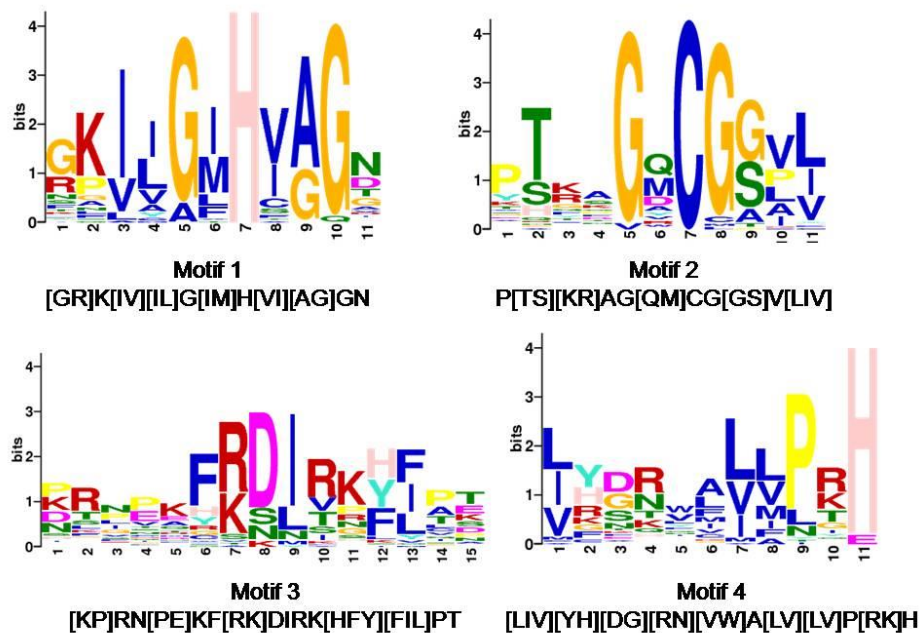


Figure 2. Conserved amino acids in selected areas by MEME.

Correlation between the conserved motifs obtained and the conserved domain detected by CDD (Conserved Domain Database). Sequences from the two most conserved motifs (motifs 1 and 2) separated by three amino acids (for example, rhinovirus 3C protease) corresponds the peptidase 3C superfamily domain (pfam00548 - 25 present in the studied viruses) when analyzed in the Conserved Domain Database (CDD), but in the case of 3C protease with larger separations between motifs 1 and 2 this domain is not detected. Instead of pfam00548 code were recognized the cl02893 domain (present in 16 of the studied viruses),

cl13774 domain (present in 7 studied viruses) or it was not possible to detect any type on 3C peptidase domain (7 studied viruses) (table 2).

Table 2. Distance between the conserved sites of the 3C-Pro in the studied viruses. * Amounts in number of amino acids. The superscripts in the names of viruses correspond to domain 3C from each according to the database of conserved domains (CDD): the domain ^a pfam00548; ^b cl02893; ^c cl13774; ^d no domain 3C protease detected.

| Virus | Between Motif 1-2* | Between Motif 2-3* | Between Motif 3-4* |
|--|-------------------------------|-------------------------------|-------------------------------|
| <i>PATTERN 3/48/35</i> | | | |
| Avian sapelovirus ^a | 3 | 48 | 35 |
| Pigeon picornavirus B ^a | 3 | 48 | 35 |
| <i>PATTERN 3/48/36</i> | | | |
| Bat picornavirus 1 ^a | 3 | 48 | 36 |
| Bat picornavirus 2 ^a | 3 | 48 | 36 |
| Bat picornavirus 3 ^a | 3 | 48 | 36 |
| Canine picornavirus ^a | 3 | 48 | 36 |
| Feline picornavirus ^a | 3 | 48 | 36 |
| Simian sapelovirus 1 ^a | 3 | 48 | 36 |
| <i>PATTERN 3/48/37</i> | | | |
| Enterovirus A ^a | 3 | 48 | 37 |
| Enterovirus B ^a | 3 | 48 | 37 |
| Echovirus E6 ^a | 3 | 48 | 37 |
| Human coxsackievirus A24 ^a | 3 | 48 | 37 |
| Human coxsackievirus B4 ^a | 3 | 48 | 37 |
| Human enterovirus C109 ^a | 3 | 48 | 37 |
| Human poliovirus1 ^a | 3 | 48 | 37 |
| Human poliovirus 2 ^a | 3 | 48 | 37 |
| Rhinovirus A ^a | 3 | 48 | 37 |
| Rhinovirus C ^a | 3 | 48 | 37 |
| Simian enterovirus 46 ^a | 3 | 48 | 37 |
| <i>PATTERN 7/45/34-35</i> | | | |
| Aichivirus A ^b | 7 | 45 | 34 |
| Aichivirus B ^b | 7 | 45 | 35 |
| <i>PATTERN 7/46/21-22</i> | | | |
| Ljungan virus ^a | 7 | 46 | 22 |
| Human parechovirus 1 ^a | 7 | 46 | 21 |
| Human parechovirus 6 ^a | 7 | 46 | 21 |
| <i>PATTERN 7/46/45</i> | | | |
| Swine pasivirus 1 ^b | 7 | 46 | 45 |
| <i>PATTERN 7/49-51/40-44</i> | | | |
| Cosavirus A ^b | 7 | 49 | 40 |
| Human cosavirus E1 ^a | 7 | 49 | 40 |
| Encephalomyocarditis virus ^b | 7 | 51 | 40 |
| Theiler's encephalomyelitis virus ^a | 7 | 51 | 40 |
| Foot-and-mouth disease virus type O ^a | 7 | 51 | 44 |
| <i>PATTERN 7/50/43</i> | | | |

| | | | |
|---|---|----|----|
| Equine rhinitis B virus 2 ^a | 7 | 50 | 43 |
| <i>PATTERN 8/60/44-46</i> | | | |
| Avian encephalomyelitis virus ^b | 8 | 60 | 44 |
| Hepatitis A Virus ^b | 8 | 60 | 46 |
| <i>PATTERN 3/56/44</i> | | | |
| Ectropis obliqua picorna-like virus ^d | 3 | 56 | 44 |
| Perina nuda virus ^d | 3 | 56 | 44 |
| <i>PATTERN 4/69/48-52</i> | | | |
| Spodoptera exigua iflavivirus ^d | 4 | 69 | 48 |
| Infectious flacherie virus ^d | 4 | 69 | 52 |
| <i>PATTERN 6/51/64</i> | | | |
| Deformed wing virus ^c | 6 | 51 | 64 |
| Kakugo virus ^c | 6 | 51 | 64 |
| Varroa destructor virus-1 ^c | 6 | 51 | 64 |
| <i>PATTERN 6/53-55/46-48</i> | | | |
| Nilaparvata lugens honeydew virus-1 ^d | 6 | 53 | 46 |
| Brevicoryne brassicae picorna-like virus ^b | 6 | 55 | 48 |
| <i>PATTERN 6/54/66</i> | | | |
| Formica exsecta virus | 6 | 54 | 66 |
| <i>PATTERN 6-7/57-58/48-50</i> | | | |
| Halyomorpha halys virus ^d | 6 | 57 | 50 |
| Lygus lineolaris virus 1 ^b | 6 | 57 | 48 |
| Sacbrood virus ^b | 7 | 58 | 50 |
| <i>PATTERN 9/56/56</i> | | | |
| Slow bee paralysis virus ^b | 9 | 56 | 56 |
| <i>PATTERN 8/63/49</i> | | | |
| Acute bee paralysis virus ^c | 8 | 63 | 49 |
| Israeli acute paralysis virus ^c | 8 | 63 | 49 |
| Kashmir bee virus ^b | 8 | 63 | 49 |
| <i>PATTERN 8/68/68</i> | | | |
| Drosophila C virus ^b | 8 | 68 | 68 |
| <i>PATTERN 8/69/47</i> | | | |
| Taura syndrome virus ^c | 8 | 69 | 47 |
| <i>PATTERN 9/66/61</i> | | | |
| Black queen cell virus ^b | 9 | 66 | 61 |
| <i>PATTERN 6/49/46</i> | | | |
| Rhizosolenia setigera RNA virus 01 ^b | 6 | 49 | 46 |

Analysis of 3C protease regarding the distances between conserved motifs. Analysis of the inter- motifs regions size allowed to obtain at least 21 patterns (summarized in table 2). A comparison of this patterns with the data obtained from the CDD allowed us to infer a lack of knowledge about the real conserved domain of the PRO 3C-protease since some families possessing 3C protease and detected by our approach have viruses which did not present a 3C-PRO domain by the CDD analysis tool . At the same time, a matrix containing the patterns of inter-distances motifs was subjected to parsimony analysis. The tree obtained clustered most viruses for the family picornaviridae, but mixing members of the families Iflaviridae and Dicistroviridae (Figure 3).

Phylogenetic analysis of the conserved motifs. After obtaining the conserved motifs they were analyzed together in a unique sequence, using only the motif 1 and 2 together, because of their proximity and conservation among studied virus, and motifs 1, 3 and 4 alone. The method chosen was the Maximum Likelihood using the evolutionary model with the highest statistical support through the use of MEGA 6. The trees based on conserved motifs had lower statistical power than that obtained with conserved motifs as RdRp or S3H. The tree obtained using motifs 1 and 2 together (Figure 4) showed greatest statistical support and better clustering of families than the trees obtained from isolated motif 1, motif 3, motif 4 , or all motifs together (Trees not shown). Phylogeny of the motif 1 and 2 formed a large clade of member of the Picornaviridae family (containing 23 of the 33 Picornavirus included in this study), a clade of Dicistroviridae family members (containing 6 of the 7 viruses analyzed) and Iflaviridae clade (containing 11 of the 14 analyzed virus). Some of these virus families eventually split into isolated groups or forming virus clades from different families and with acceptable statistical support as the case of the combination between Spodoptera and Taura virus (Iflavirus and Dicistrovirus, respectively). In motif 3 tree, there was a greater separation of families , which produced many groups. The motif 4 tree, in turn, established a large group of the Picornaviridae family and several isolated groups with Iflaviridae, Dicistroviridae and also Picornaviridae. Some small clades, for example, those of KIA complex (Kashmir bee virus, Israeli acute paralysis virus and Acute bee paralysis virus), Deforming Wing Virus, Kakugo Virus and Varroa destructor Virus and Ectropis and Perina viruses remained grouped independently of the analyzed system, showing high conservation between these small viral groups.

Comparison between data obtained and the phylogenetic data based on the sequence of the RdRp polymerase. As polymerase RdRp is one of the best sequences to evaluate phylogenetically Class IV viruses, we analyzed this sequence by Maximum Likelihood method on all viruses of this work, the trees produced showed good statistical support (Figure 5) and showed the expected separation of Picornaviridae, Dicistroviridae and Iflaviridae. Comparing the RdRp tree data with the data mentioned above was possible to observe:

- a) All viruses showing pattern 3/48/37, were recognized with the domain pfam00548 and formed a clade either in RdRp and 3C-PRO, the difference between the two trees is a change in the clade location of Enterovirus and the presence of Avian Sapelovirus in 3C-pro tree that is not present in the same clade of RdRp tree;
- b) All viruses showing the pattern 3/48/35-36 were also recognized with the domain pfam00548 and formed a clade both in RdRp as in 3C-PRO trees, the difference between the trees is a change of placement for some viruses and the presence of the Avian Sapelovirus in the RdRp tree;
- c) Viruses with 7/51/44 pattern (pfam00548), 7/51/40 (pfam00548 in one virus and cl02893 in one virus), 7/49/40 (cl02893) and 8/60/44-46 (pfam00548) form a clade in the 3C PRO tree including one iflavirus (Slow Bee Paralysis Virus, 9/56/56 , cl02893). In the RdRp tree the iflavirus is not present, and includes a clustering among Aichivirus A and B (7/45/34-35, cl02893) and Equine rhinitis viruses (7/50/43, pfam00548) . In this case, it should be emphasized the lack of phylogenetic logic in the CDD data because pfam00548 and cl02893 were found , in members of a same clade, can be mentioned the clade form by Encephalomyocarditis virus and Theiler 's Encephalomyelitis virus that are grouped in both of trees, share the same pattern of distance between conserved motifs, and however have different CDD domains. Moreover this clade is related to the aforementioned picornaviridae groups when evaluated in RdRp tree, while in the 3C-PRO tree are related to the Iflavirus clade.
- d) The most isolated clade within the family Picornaviridae, using as a template the RdRp tree, is formed by viruses with pattern 7/46/21-22-45 (pfam00548 for 7/46/21-22 and cl02893 for 7/46/ 5) and are also grouped in the 3C-PRO tree. However, in the 3C-PRO tree this clade is more related to the group with pattern 3/48/35-36-37.

- e) Clades of iflavivirus and dicistrovirus are clustered in the RdRp tree, each family forming a different sub-clade. In the 3C-PRO tree the iflavivirus share a clade with a group of picornavirus while the dicistrovirus group remains separate from this. The viruses within each group presented somewhat differences on the type of associations. When comparing in a parsimonious way, the distance patterns within these families were more related in comparison to the RdRp tree than the 3C tree.
- f) Considering the above result, we constructed a matrix with the pattern of distance between the conserved motifs and analyze them by the parsimony method. The tree obtained (Figure 3) showed great similarity to the RdRp tree. Highlighting the group of picornavirus with pattern 3/48/35-36-37 clustering near and the separating Picornaviridae members with pattern 7/45/34-35 and 7/46/21-22 from other picornavirus. The cluster formed by the Hepatitis A Virus and Avian Encephalomyelitis Virus, which is farthest from the major clade of the picornavirus in RdRp tree also departs from this tree, being separated from the large group of picornavirus by a set of Iflavivirus. Finally, are some dicistrovirus and iflavivirus. Interestingly, the clade where Dicistrovirus and Iflavivirus are mixed correspond mostly to viruses that infect bees.

DISCUSSION

Viral infection create a great charge for any eukaryote. As a result, they have efficient mechanisms to combat viral infections that differ in vertebrates, insects and plants (3). The Picornaviridae family has a broad host range, being a parasite present in all kingdoms. Its viral genome consisting of single-stranded positive sense RNA that is translated into a single polyprotein, which is subsequently processed into individual proteins through its autocleavage by viral proteases (1). In the case of picornaviridae, 2A protease family catalyzes the first cleavage between structural and non-structural proteins, and subsequently 3C protease generates the functional proteins and enzymes (20, 21); however in other families of this order who do not coding a 2A protease this task is performed by the 3C-PRO protease. Baker and Schroeder (6) showed that RdRp is a valid marker to establish a taxonomy using picornaviruses sequences and also for their identification. Although the RdRp polymerase is the only shared protein in Picorna-like superfamily, the evolution of viruses belonging to it

cannot be reduced to this single protein (13), and, on that basis, we analyze the evolution of 3C - Pro protease. The 3C-Pro region encodes a protease that plays a key role in the cleavage of viral polyprotein (21). In the case of picornaviruses and other Picorna-like viruses the 3C-Pro is also responsible for cleavage of host proteins to get rid of intracellular antiviral mechanisms (21, 23,20), which provides an important evolutionary pressure in this region, because mutation in the recognition sites of the protein sequence should be able to cleave the polyprotein, at the same time, cleaving antivirals host proteins. Both actions are very important for the successful of the viral infection. In this study, we present a way to establish the relationship of 3C-Pro from Picornaviridae, Dicistroviridae and Iflaviridae family. From the analysis of their conserved motifs, to avoid the noise provided by the poorly conserved inter-motif regions, because according to Koonin (12) to find conserved regions within viral genomes is the first step for comparative studies of these genomes.

Due to the lack of sequence similarity of 3C-Pro with the 3C-like proteases it has been suggested that they have no evolutionary relationship with cysteine proteases, and its analogy could be explained by convergence (9); however, the relationship found between the patterns of inter-motifs distance and phylogeny of RdRp polymerase could also be explained a large number of mutations in the 3C-pro sequence. We believe that the great RdRp conservation between picornavirus is a function of the viral replication, because the enzyme reaction remains equal from the standpoint of the substrates , except for the sequence recognized by VPg protein (necessary for the initiation of replication), while 3C-Pro requires modify its strategy cleavage every time your host changes its defense strategy or when these viruses infect a new organisms. Thus, the 3C-pro region have evolved independently from RdRp in Picorna-like superfamily and picornaviruses, due to be parasites of a broad host range and have to get rid of the various defense mechanisms.

The best described picornavirus pathogens are Poliovirus (PV), human rhinovirus (HRV) , foot- and- mouth disease virus (FMDV) and hepatitis A (HAV) (19) . Based on the motif 1 and 2 , these viruses exhibited a proximity, which contributes to stimulate the use of this region to establish phylogenetic relationships. Groups such as the iflavírus DWV, VDV and KV, which remained in the same group in all the tested system used in this study. Its high degree of similarity can be explained by the fact of possessing a possible common ancestor (6) or further by the 3C protease sites of DWV and VDV remain identical because they are parasites at the same host (14) , the same sequence of event may also be applied to

dicistrovirus KIA complex (Kashmir bee virus, Israeli acute paralysis virus, Acute bee paralysis virus) which also share the same host (Bees) or Perina nuda virus and Spodoptera exigua virus, viruses that infect other species, such Lepidoptera or moth and butterflies, in addition to several clades of viruses that infect humans .

In the phylogenetic significance, 3C protease is an attractive target for the development of antiviral drugs (21, 20, 16). In these studies, several non-structural proteins aim the identification of conserved region to be used as a target for binding of antiviral substances (19). Thus, as the better result were obtained with the motif 1 and 2, this region may serve as a marker for targeting of antiviral drugs against the 3C protease and also in establishing phylogenetic relationships or to detect the 3C proteinase domain in the description of new viruses. Therefore, considering the lack of consistency phylogenetic in the use of pfam00548, cl13774 and cl02893 domain, and the absence of 3C domain recognition in some viruses of the Picorna-like superfamily from CDD tool, we propose as 3C domain the sequence formed by the motif 2 separate from the motif 1 by an X_{3-9} interval. (P[TS][KR]AG[QM]CG[GS]V[LIV] X_{3-9} [GR]K[IV][IL]G[IM]H[VI][AG]GN).

CONCLUSION

The use of conserved motifs can be a good alternative to conduct phylogenetic studies of highly variable sequences. In the case of 3C - PRO from Picorna-like superfamily were recognized four conserved motifs that were present in all viruses in the same order , with the exception of Aphid lethal paralysis virus that did not show the motif 3 .

The use of domains 1 and 2 provided the tree with greater robustness displayed until today for this viral protein. Because the RdRp has no proofreading system is expected to have a large number of errors, single nucleotide mutations during replication , but would be less likely to changes structurally therefore, it would be easier to maintain the structure between the domains than the aminoacids sequences. In this case the evolutionary analysis of intermotifs regions can be a useful and valuable tool .

The weigh matrix from motif 2 and motif 1 separated by a X_{3-9} sequence could be used as 3C protease domain for study picornavirus and related viruses.

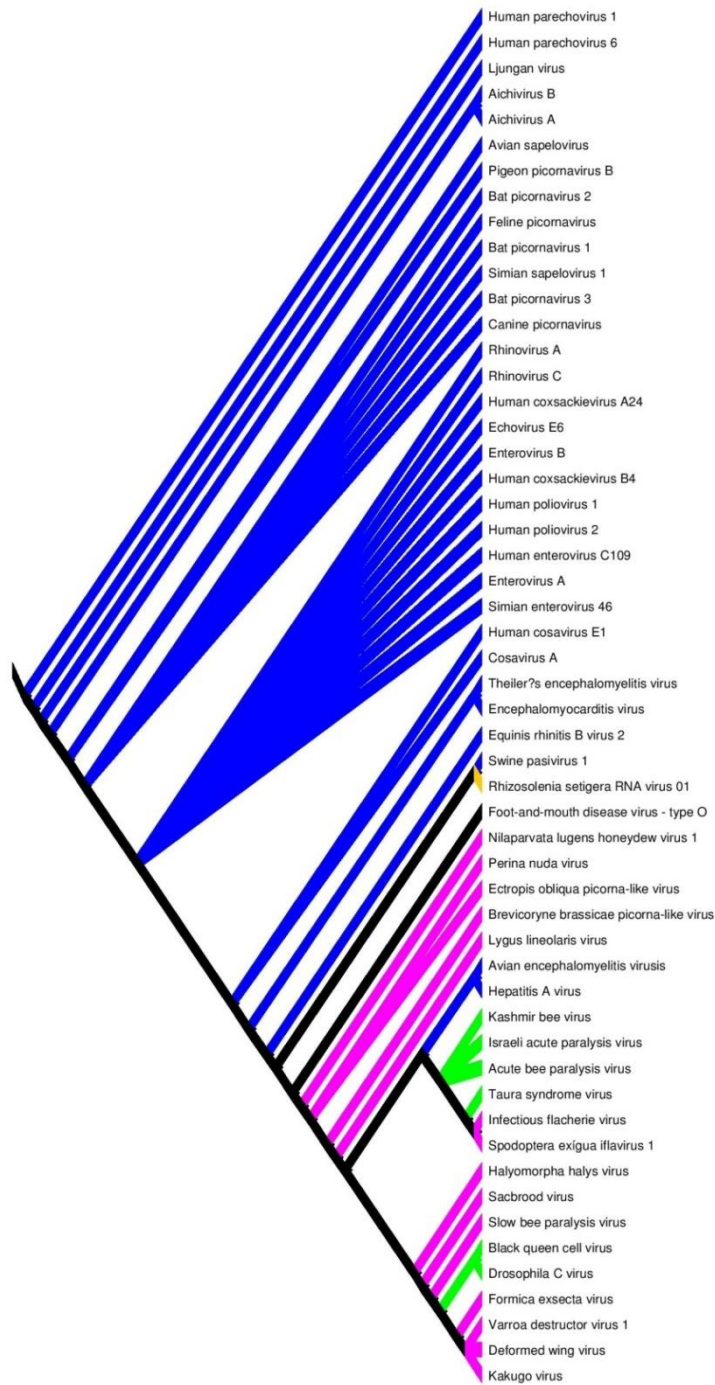


Figure 3. Phylogenetic analysis by parsimony and the distances between the conserved sites. In blue are represented the virus of family Picornaviridae, are represented in pink viruses of family Iflaviridae, in green the family of Dicistroviridae viruses and the virus that orange was chosen for outgroup in the analysis by maximum likelihood.

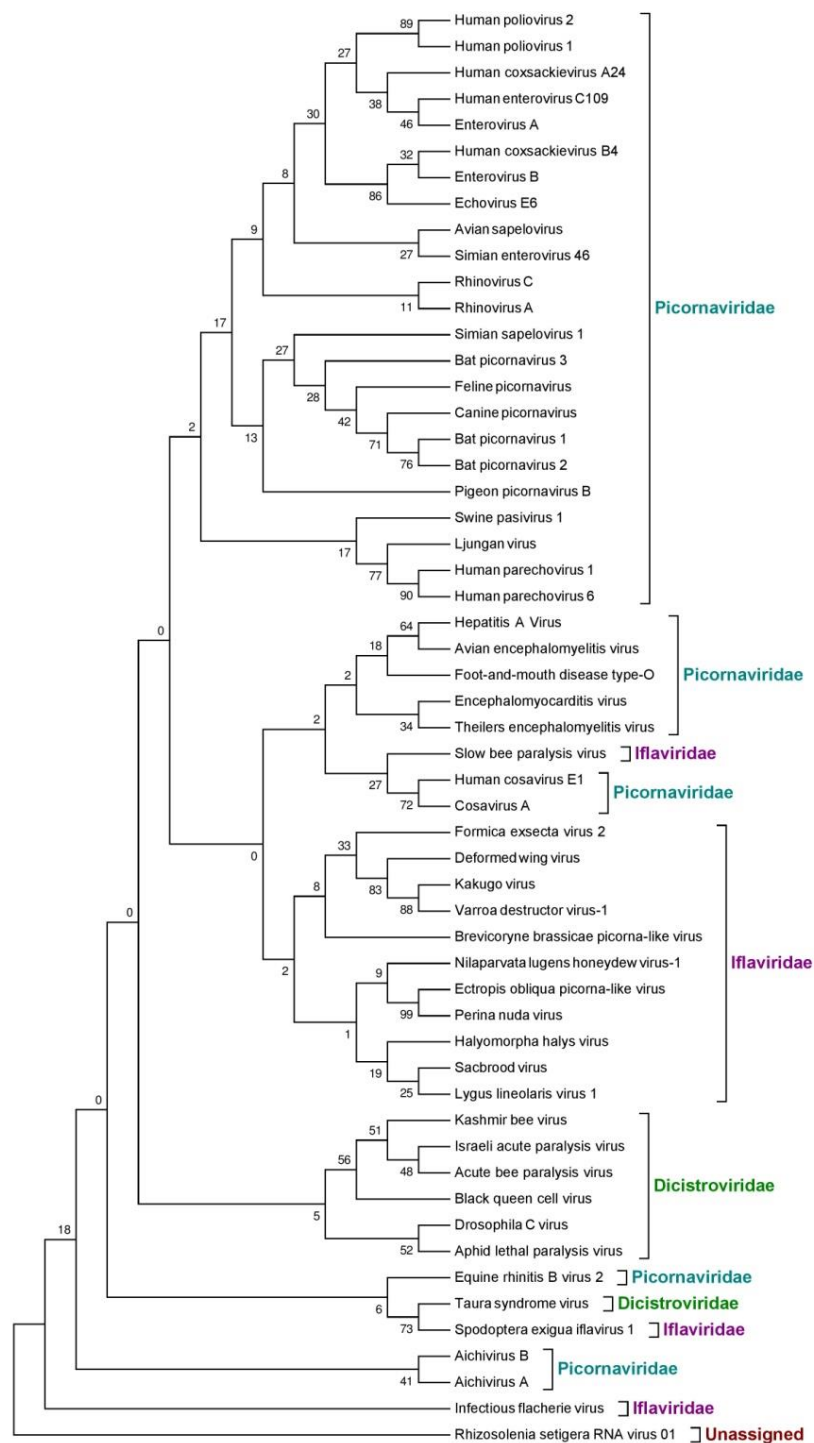


Figure 4. Molecular Phylogenetic analysis by Maximum Likelihood method of motifs 1 and 2. The evolutionary history was inferred by using the Maximum Likelihood method. The bootstrap consensus tree inferred from 5000 replicates is taken to represent the evolutionary history of the taxa analyzed. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.6718)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 5.9550% sites). The analysis involved 55 amino acid sequences. All positions containing gaps and missing data were eliminated.

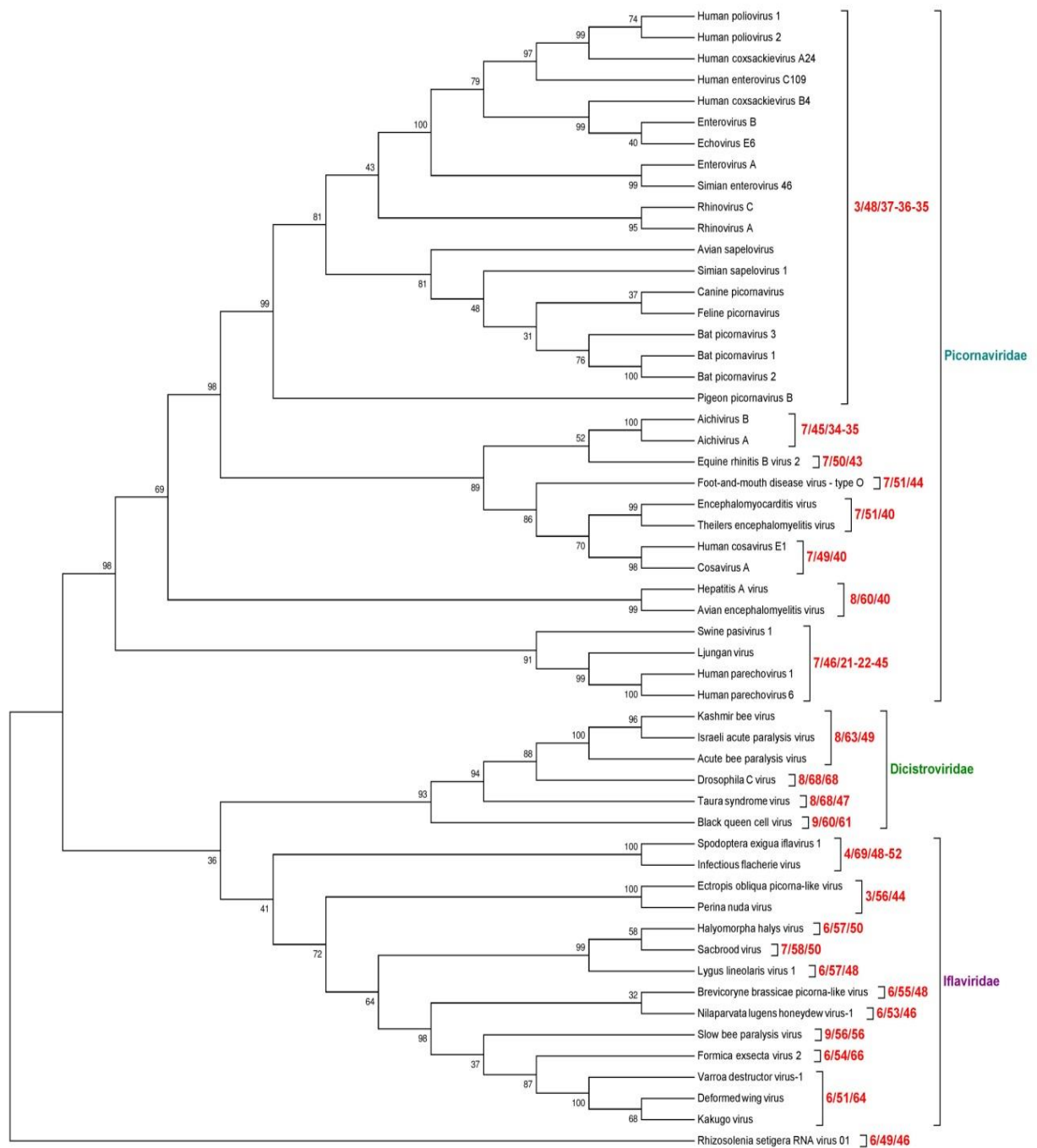


Figure 5. The evolutionary history was inferred by using the Maximum Likelihood method to RdRp with the standard length of the 3C-Pro motifs (red) and families.. The bootstrap consensus tree inferred from 1000 replicates. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.7188)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 6.4258% sites). The analysis involved 54 amino acid sequences. All positions containing gaps and missing data were eliminated.

REFERENCES

- 1 AMINEVA, S. P., AMINEV, A. G., PALMENBERG, A. C., GERN J. E. Rhinovirus 3C protease precursors 3CD and 3CD' localize to the nuclei of infected cells. *Journal of General Virology*, 85: 2969-2979, 2004.
- 2 ARNOLD, E., LUO, M., VRIEND, G., ROSSMANN, M. G., PALMENBERG, A. C., PARKS, G. D., NICKLIN, M. J. H., WIMMER, E. Implications of the picornavirus capsid structure for polyprotein processing. *Proc. Natl. Acad. Sci. USA*, 84: 21-25, 1987.
- 3 AZZAMI, K., RITTER, W., TAUTZ, J.BEIER, H. Infection of honey bees with acute bee paralysis virus does not trigger humoral or cellular immune responses. *Arch Virol*, 157(4):689-702, 2012.
- 4 BAILEY, T. L. AND GRIBSKOV, M. "Combining evidence using p-values: application to sequence homology searches", *Bioinformatics*, 14(1):48-54,1998.
- 5 BAILEY,T.L.,BODÉN, M., BUSKE, F. A.,FRITH, M. , GRANT,C. E., CLEMENTI, L., REN,J., LI, W.W., NOBLE, W.S. "MEME SUITE: tools for motif discovery and searching", *Nucleic Acids Research*, 37:W202-W208, 2009.
- 6 BAKER, A. C. and SCHROEDER, D. C. The use of RNA-dependent RNA polymerase for the taxonomic assignment of Picorna-like viruses (order Picornavirales) infecting *Apis mellifera* L. populations. *Virology Journal*, 5:10, 2008.
- 7 BALTIMORE, D. .Expression of animal virus genomes. *Bacteriol Rev* 35 (3): 235–41,1971.
- 8 FRANCKI, R. I. B.; FAUKUET, C. N., KNUDSON, D. L.; BROWN, F. Classification and nomenclature of viruses. Fifth Report of the international committee on taxonomy of viruses. *Archives of Virology (Suppl. 2)*, 1, 1991.
- 9 GORBALENYA, A. E., DONCHENKO, A. P., BLINOV, V. M., KOONIN, E. V. Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases: A distinct protein superfamily with a common structural fold. *FEB*, 243(2):103-114, 1989.
- 10 GOVAN, V.A., LEAT, N., ALLSOPP, M., DAVISON, S. Analysis of the complete genome sequence of Acute bee paralysis virus shows that it belongs to the novel group of insect-infectin RNA viruses. *Virology*, 277:457-463, 2000.
- 11 HANCAK R, SEMLER B. L., ARIGA, H. , ANDERSON, C.W. , WIMMER, E. Expression of a cloned gene segment of poliovirus in *E. coli*: evidence for autocatalytic production of the viral proteinase. *Cell*. 37 (3):1063–1073, 1984.
- 12 KOONIN, E. V.; DOLJA, W. Evolution and taxonomy of positive strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol Biol* 28: 375-430, 1993.
- 13 KOONIN, E. V.; WOLF, Y. I., NAGASAKI, K., DOLJA, V. V. The Big Bang of Picorna-like virus evolution antedates the radiation of eucariotic supergroups. *Nature Reviews - Microbiology*, vol. 6, p. 925-939, 2008.
- 14 LANZI, G. et al.. Molecular and biological characterization of Deformed Wing Virus of

honeybees (*Apis mellifera* L.). *Journal of Virology*, 80: 4998-5009, 2006.

15 MARCHLER-BAUER A, ZHENG C, CHITSAZ F, DERBYSHIRE MK, GEER LY, GEER RC, GONZALES NR, GWADZ M, HURWITZ DI, LANCZYCKI CJ, LU F, LU S, MARCHLER GH, SONG JS, THANKI N, YAMASHITA RA, ZHANG D, BRYANT SH. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* 41: 348-352, 2013.

16 MILLS, J. S. Virus proteinase inhibitors What next after HIV?. *Antiviral Chemistry & Chemotherapy.* 7(6): 281-293, 1996.

17 NAKASHIMA, N. and NAKAMURA, Y. Cleavage sites of the “P3 region” in the structural polyprotein precursor of a dicistrovirus. *Archives of Virology*, 153: 1955-1960, 2008.

18 National Center of Biotechnology Institute: www.ncbi.nlm.nih.gov/

19 NORDER, H. et al.. Picornavirus non-structural proteins as targets for new anti-virals with broad activity. *Antiviral Research*, 89:204-218, 2011.

20 PALMENBERG, A. C. Proteolytic processing of picornaviral polyprotein. *Annu. Rev. Microbiol.* 44:603-623, 1990.

21 PORTER, A.G. Picornavirus nonstructural proteins: emerging roles in virus replication and inhibition of host cell functions. *Journal of Virology* 67(12): 6917-6921, 1993.

22 TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. and KUMAR, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* 30: 2725-2729, 2013.

23 WANG A, Q. M. and CHEN, S. Human Rhinovirus 3C Protease as a Potential target for the development of antiviral agents. *Current Protein and Peptide Science*, 8, 19-27, 2007.

24 WANG, H. and LIANG, P. Picornaviral 3C protease inhibitors and the dual 3C protease/coronaviral 3C-like protease inhibitors. *Expert Opinion*, 20: 509-71, 2010.

Supplementary material 1

Motif 1 position-specific scoring matrix

log-odds matrix: length= 20 w= 11 n= 7675 bayes= 7.11422 E= 5.8e-235

| | | | | | | | | | | | | | | | | | | | |
|------|------|------|------|------|------|------|-------|------|------|-------|------|------|------|------|------|------|-------|------|------|
| -102 | -461 | -278 | -190 | 30 | 222 | -307 | -189 | -30 | -387 | -341 | 48 | -376 | -6 | 202 | 8 | -71 | -198 | -358 | -116 |
| 8 | -487 | -295 | -213 | -25 | -71 | -325 | -474 | 323 | -119 | -371 | -37 | 172 | -11 | -237 | -111 | -320 | -435 | -383 | -426 |
| -172 | -439 | -540 | -512 | -384 | -621 | -598 | 328 | -500 | -43 | -228 | -516 | -570 | -458 | -518 | -456 | -412 | 128 | -418 | -475 |
| 42 | 64 | -607 | -512 | -343 | -565 | -505 | 228 | -508 | 182 | -42 | -517 | -530 | -416 | -478 | -406 | -375 | 84 | -344 | 36 |
| 84 | -511 | -482 | -522 | -640 | 335 | -584 | -680 | -534 | -668 | -597 | -464 | -563 | -505 | -508 | -385 | -528 | -605 | -493 | -639 |
| -564 | -588 | -878 | -784 | 142 | -844 | -753 | 252 | -788 | 51 | 349 | -804 | -749 | -617 | -728 | -718 | -184 | -188 | -476 | -607 |
| -903 | -858 | -870 | -950 | -898 | -840 | 486 | -1076 | -949 | -945 | -1035 | -929 | -825 | -854 | -869 | -901 | -947 | -1048 | -745 | -903 |
| -129 | 193 | -630 | -539 | -13 | -584 | -538 | 198 | -537 | -315 | -306 | -544 | -554 | -448 | -509 | -5 | -164 | 252 | -382 | -462 |
| 361 | -508 | -771 | -763 | -756 | 190 | -779 | -801 | -803 | -761 | -712 | -693 | -631 | -644 | -722 | -414 | -533 | -646 | -668 | -813 |
| -408 | -597 | -505 | -550 | -676 | 346 | -610 | -723 | -560 | -709 | -643 | -493 | -610 | 27 | -535 | -460 | -613 | -674 | -518 | -666 |
| 10 | -455 | 155 | -78 | -116 | 42 | -310 | -101 | -204 | -380 | -60 | 260 | -378 | -147 | -16 | -105 | 86 | -115 | -355 | -399 |

Motif 2 position-specific scoring matrix

log-odds matrix: length= 20 w= 11 n= 7675 bayes= 7.11422 E= 1.6e-191

| | | | | | | | | | | | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| -36 | -44 | -114 | -15 | -118 | -125 | -308 | -455 | 45 | -199 | -350 | -111 | 317 | -6 | -236 | -38 | -25 | -119 | 110 | 124 |
| -148 | -401 | -467 | -486 | -509 | -241 | 111 | -441 | -437 | -509 | -386 | -138 | -507 | 24 | -142 | 171 | 322 | -409 | -433 | -568 |
| -35 | 54 | -288 | -81 | 71 | 21 | -79 | -99 | 196 | -191 | -59 | -279 | -382 | 145 | 202 | -238 | -23 | -192 | -351 | 30 |
| 188 | -469 | 47 | -12 | 70 | -206 | -302 | -455 | 95 | -199 | -62 | 13 | 57 | 146 | -15 | 113 | -23 | -201 | -360 | -400 |
| -364 | -557 | -473 | -517 | -644 | 345 | -580 | -680 | -527 | -675 | -606 | -460 | -579 | -509 | -503 | -420 | -571 | -139 | -489 | -636 |
| 44 | -425 | 120 | -236 | -390 | -446 | -340 | -366 | -105 | -182 | 302 | -311 | -406 | 340 | 30 | -265 | -312 | -57 | 211 | 31 |
| -436 | 544 | -615 | -562 | -572 | -659 | -613 | -556 | -630 | -568 | -517 | -605 | -616 | -540 | -558 | -529 | -513 | -604 | -548 | -660 |
| -266 | 10 | -367 | -410 | -546 | 339 | -479 | -580 | -420 | -266 | -117 | -353 | -487 | -407 | -399 | -320 | -478 | -533 | -391 | -536 |
| 107 | -448 | -728 | -710 | -710 | 242 | -740 | -745 | -756 | -711 | -653 | -634 | -128 | -591 | -677 | 269 | -80 | -584 | -637 | -773 |
| 136 | -360 | -581 | -486 | -95 | -538 | -477 | 65 | -160 | 95 | -254 | -490 | 190 | -389 | -451 | -378 | -351 | 207 | -320 | -402 |
| -326 | -31 | -611 | -515 | -95 | -570 | -95 | 168 | -511 | 243 | -256 | -521 | -531 | -413 | -478 | -411 | -382 | 136 | -335 | -111 |

Motif 3 position-specific scoring matrix

log-odds matrix: length= 20 w= 15 n= 7455 bayes= 7.07195 E= 5.6e-169

| | | | | | | | | | | | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| -102 | -470 | 170 | -185 | -435 | -415 | 59 | -104 | 185 | -397 | -348 | 156 | 211 | -140 | -86 | 8 | -143 | -201 | -362 | -402 |
| -227 | -468 | -112 | -184 | 30 | -124 | -78 | -192 | 13 | -29 | -62 | -109 | -122 | -4 | 298 | 70 | 104 | -412 | 111 | -400 |
| -116 | -28 | -561 | -465 | 43 | 23 | -457 | 64 | -460 | 43 | -37 | 185 | 174 | -369 | -36 | -4 | -152 | -3 | 119 | -384 |
| -228 | 110 | -112 | 203 | -26 | -124 | -303 | -104 | -195 | -393 | -345 | 12 | 242 | 65 | -15 | 8 | -23 | 3 | -359 | -400 |
| 96 | -43 | 11 | 113 | -433 | -414 | -302 | -455 | 232 | -118 | 82 | -36 | -122 | 174 | 32 | -103 | -23 | -201 | -360 | -116 |
| -454 | 116 | -612 | -561 | 378 | -77 | 157 | -463 | -565 | -411 | -423 | -153 | -594 | -471 | 17 | -159 | -529 | -206 | 125 | 131 |
| -150 | -630 | -612 | -435 | -705 | -620 | -443 | -635 | 252 | -555 | -526 | -473 | -601 | -254 | 359 | -5 | -178 | -216 | -505 | -597 |
| -473 | -640 | 374 | -385 | -689 | -451 | -437 | -798 | -6 | -737 | -732 | 116 | -559 | -382 | -514 | 119 | -478 | -757 | -627 | -604 |
| -561 | -588 | -845 | -763 | -371 | -829 | -726 | 323 | -762 | 117 | 44 | 45 | -737 | -602 | -709 | -698 | -612 | -397 | -457 | -117 |
| -120 | -368 | -536 | -433 | -335 | -524 | -446 | 10 | -58 | -156 | -263 | -455 | -491 | -331 | 315 | 65 | 123 | 97 | -321 | -402 |
| -104 | -477 | -281 | -192 | -443 | -4 | 6 | -192 | 287 | -403 | -355 | 78 | 114 | -5 | 98 | 7 | -145 | -422 | -367 | -409 |
| -450 | -503 | -577 | -532 | 281 | -647 | 306 | -195 | -71 | -119 | -460 | -535 | -577 | -460 | -506 | -156 | -552 | -525 | -194 | 289 |
| -486 | -531 | -782 | -685 | 305 | -740 | -654 | 208 | -683 | 141 | -40 | -695 | -676 | -57 | -636 | -594 | -541 | -105 | -425 | 37 |
| 152 | 159 | -530 | -437 | -316 | -129 | -448 | -22 | -435 | -9 | -245 | -456 | 258 | 95 | -125 | -56 | 106 | -38 | -305 | -386 |
| 11 | -469 | 100 | 165 | -434 | -414 | -302 | -455 | 115 | -199 | -347 | 49 | 57 | -138 | -86 | 94 | 170 | -414 | -360 | -116 |

Motif 4 position-specific scoring matrix

log-odds matrix: length= 20 w= 11 n= 7675 bayes= 7.11422 E= 3.1e-132

| | | | | | | | | | | | | | | | | | | | |
|------|------|------|------|------|------|------|-------|------|------|------|------|------|------|------|------|------|-------|------|------|
| -152 | -540 | -814 | -728 | -102 | -787 | -723 | 209 | -731 | 210 | 42 | -741 | -718 | -602 | -693 | -647 | -561 | 161 | -486 | -597 |
| -143 | 113 | -491 | -405 | 123 | -7 | 253 | -489 | 66 | -427 | -416 | -441 | -512 | -296 | 159 | -391 | -461 | -478 | -231 | 286 |
| -38 | -484 | 237 | -16 | -25 | 130 | -315 | -475 | -33 | -415 | -367 | 80 | -387 | -9 | -248 | 148 | -73 | -433 | -378 | -116 |
| -235 | -475 | -113 | -190 | -118 | -207 | 59 | -462 | 96 | -402 | -353 | 196 | -379 | 146 | 278 | -234 | 65 | -420 | 111 | -116 |
| -116 | -29 | -154 | -141 | 81 | -219 | -452 | -21 | -447 | 19 | 44 | 2 | -482 | 95 | -423 | -56 | -74 | 125 | 458 | 75 |
| 223 | 159 | -564 | -467 | 158 | -129 | -457 | 41 | -68 | 109 | 164 | -471 | -486 | -370 | -432 | -359 | -334 | -3 | -301 | -108 |
| -153 | -547 | -844 | -773 | -457 | -839 | -809 | 137 | -786 | 252 | 39 | -789 | -764 | -672 | -764 | -711 | -579 | 172 | -560 | -659 |
| 39 | -517 | -771 | -677 | 116 | -733 | -662 | 40 | -677 | 203 | 233 | -688 | -673 | -552 | -636 | -586 | -527 | 160 | -443 | -552 |
| -261 | -486 | -440 | -393 | -500 | -503 | -472 | -213 | -400 | 9 | -471 | 22 | 408 | -325 | -407 | -320 | -184 | -475 | -476 | -144 |
| -106 | -479 | -287 | -197 | -446 | -4 | -313 | -49 | 187 | -405 | -62 | -112 | -385 | 64 | 271 | -108 | 120 | -423 | -370 | -26 |
| -854 | -841 | -815 | 13 | -841 | -824 | 478 | -1043 | -898 | -918 | -979 | -826 | -813 | -766 | -826 | -840 | -904 | -1010 | -718 | -811 |

4 CONSIDERAÇÕES FINAIS

Com o estudo filogenético da protease 3C, podemos observar a ocorrência de quatro motivos conservados. O motivo 1 e 2 mostrou um maior suporte estatístico na análise por Máxima Verossimilhança, o que nos fornece meios para estudar essa região na busca do estabelecimento de antivirais que possam combater os vírus que parasitam as abelhas.

5 REFERÊNCIAS BIBLIOGRÁFICAS

ALLEN, M. & BALL, B. *The incidence and world distribution of the honey bee viruses*. Bee World 77, 141–162, 1996.

ALLEN, W. *et al.* *The potential consequences of pollinator declines on the conservation of biodiversity and stability of food crop yields*. Conserv. Biol. 12: 8-17, 1998.

AZZAMI, K.; RITTER, W. and TAUTZ, J. *Infection of honey bees with acute bee paralysis virus does not trigger humoral or cellular immune responses*. Arch. Virol. , 2011.

BAKER, A. C. and SCHROEDER, D. C.. *The use of RNA-dependent RNA polymerase for the taxonomic assignment of Picorna-like viruses (order Picornavirales) infecting Apis mellifera L. populations*. Virology Journal, 5:10, 2008.

COX-FOSTER, D. L. *et al.*. *A metagenomic survey of microbes in honeybee Colony Collapse Disorder*. Science, 318: 283-286, 2007.

ECK, R. V. and DAYHOFF, M. O.. *Atlas of protein sequence and structure*. National Biomedical Research Foundation, Silver Spreing, maryland, 1966.

FRIES, I. e CAMAZINE, S. *Implications of horizontal and vertical pathogen transmission for honey bee epidemiology*. Apidologie 32: 199-214, 2001.

GENERSH, E. *Honey bee pathology: current threats to honey bees and beekeeping*. ApplMicrobiolBiotechnol. p. 87, 2010.

GORBALENYA, A. E., DONCHENKO, A. P., BLINOV, V. M., KOONIN, E. V.. *Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases: A distinct protein superfamily with a common structural fold*. FEB, 243(2):103-114, 1989.

HARRIS, K.S.; REDDIGARI, S. R.; NICKLIN, M.J; HAMMERLE, T.; WIMMER, E. *Purification and characterization of poliovirus polypeptide 3CD, a proteinase and a precursor for RNA polymerase*. J Virol, 66:7481-7489. 1992.

KOONIN, E. V.; DOLJA, W. Evolution and taxonomy of positive strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol Biol* 28: 375-430, 1993.

KOONIN, E. V.; WOLF, Y. I., NAGASAKI, K., DOLJA, V. V.. The Big Bang of Picorna-like virus evolution antedates the radiation of eucariotic supergroups. *Nature Reviews - Microbiology*, vol. 6, p. 925-939, 2008.

LANZI, G. et al.. Molecular and biological characterization of Deformed Wing Virus of honeybees (*Apis mellifera* L.). *Journal of Virology*, 80: 4998-5009, 2006

LIN, Jing-Yi, *et. al.* *Viral and host proteins involved in picornavirus life cycle.* *Journal of Biomedical Science*, p. 1, 2009.

MAXAM, A. M. and GILBERT, W.. *A new method for sequencing DNA.* *Proc. Natl. Acad. Sci. USA*74:560-564, 1977.

MIRANDA, J. R. et al. *The Acute bee paralysis - Kashmir bee virus - Israeli acute paralysis virus complex.* *Journal of Invertebrate Pathology*, 103: 30-47, 2010.

NAKASHIMA, N. and NAKAMURA, Y. *Cleavage sites of the "P3 region" in the nonstructural polyprotein precursor of a dicistrovirus.* *Arch. Virol.* 153: 1955-1960, 2008.

OLDROYD, B. P. *What's killing American honey bees?* *PLoS Biol.* p. 168, 2007.

NEI, M. and KUMAR, S. *Molecular Evolution and Phylogenetics.* Oxford University Press. p. 17, 2000.

PETTIS, J. *et al.* *Colony Collapse Disorder working group pathogen sub-group progress report .* *Am. Bee J.* 147: 595-597, 2007.

SCHNEIDER, H.. *Métodos de Análise Filogenética: um guia prático..* 2 ed. Ribeirão Preto: Holos Editora e Sociedade Brasileira de Genética, 2007.

SEIPELT, J., *et. al.* *The structures of picornaviral proteinases.* Virus Research. p. 159,161. 1999.

SHEN, M.; CUI, L.; OSTIGUY, N.; COX-FOSTER, D. *Intricate transmission routes and interactions between picorna-like viruses (Kashmir bee virus and Sacbrood virus) with the honeybee host and the parasitic varroa mite.* Journal of General Virology, p. 2281, 2005.

STOKSTAD, E. *The case of the empty hives.* Science 316: 970-972, 2007.

TAMURA, K; PETERSON, D; PETERSON, N; STECHER, G; NEI, M, and KUMAR, S. *MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.* Molecular Biology and Evolution 28: 2731-2739, 2011.

WANG A, Q. M. and CHEN, S.. *Human Rhinovirus 3C Protease as a Potential target for the development of antiviral agents.* Current Protein and Peptide Science, 8, 19-27, 2007.

WIESE, Helmuth. *Apicultura: Novos Tempos.* 2. ed. Guaíba: Agrolivros, p. 295, 2005.

ANEXOS

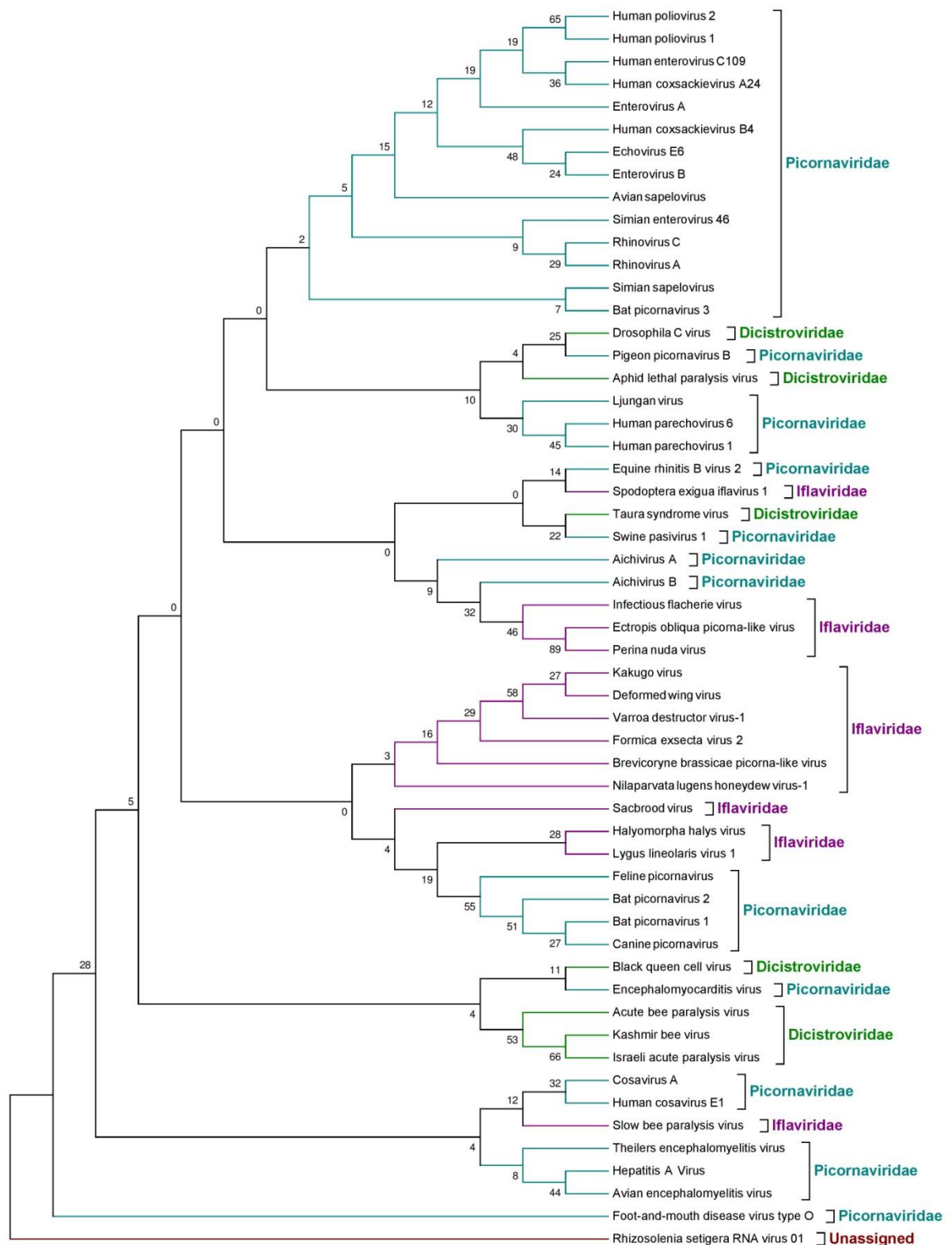


Figura 6. Árvore filogenética do motivo 1 construída pelo método de Máxima Verossimilhança.

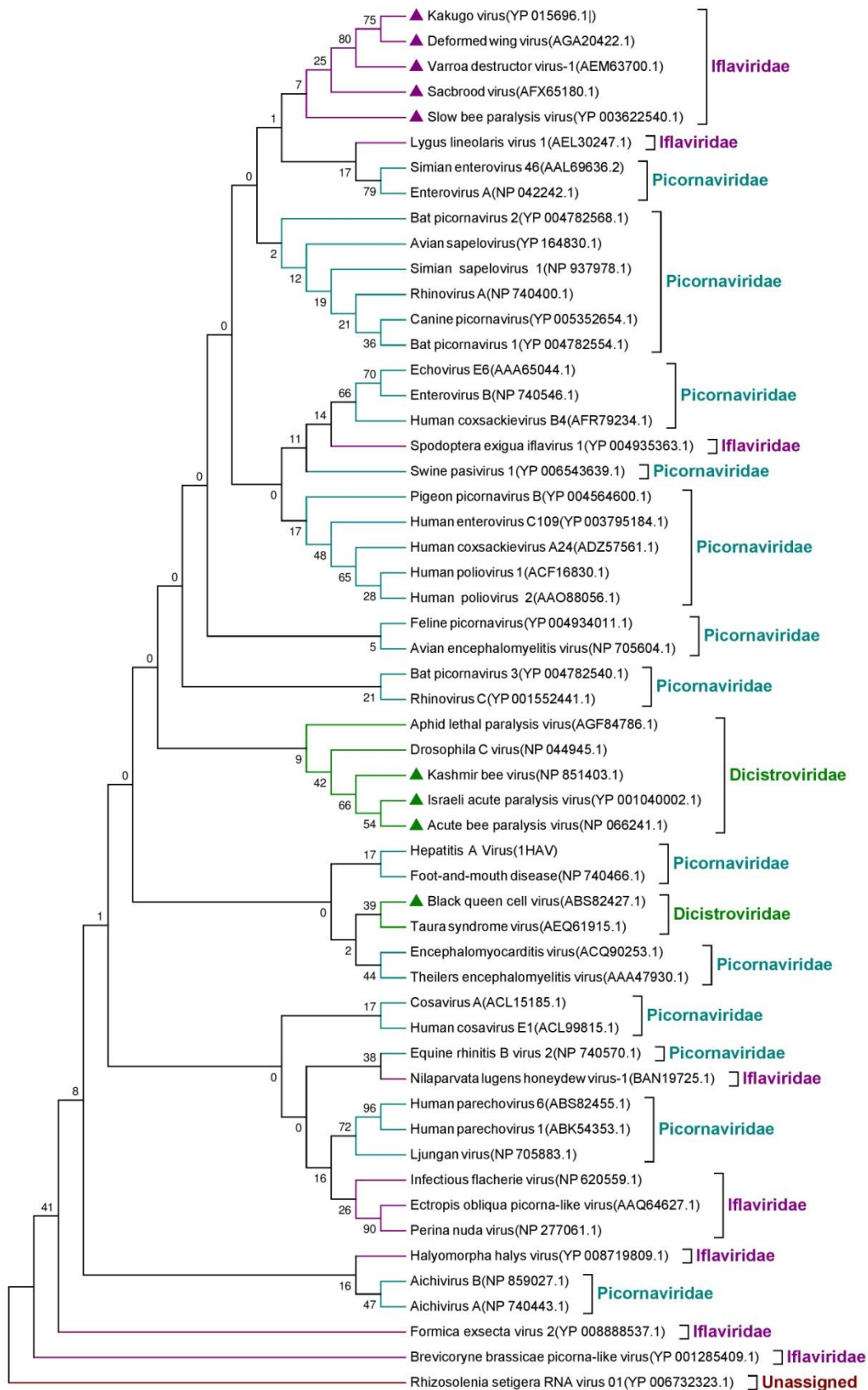


Figura 7. Árvore filogenética do motivo 3 construída pelo método de Máxima Verossimilhança.

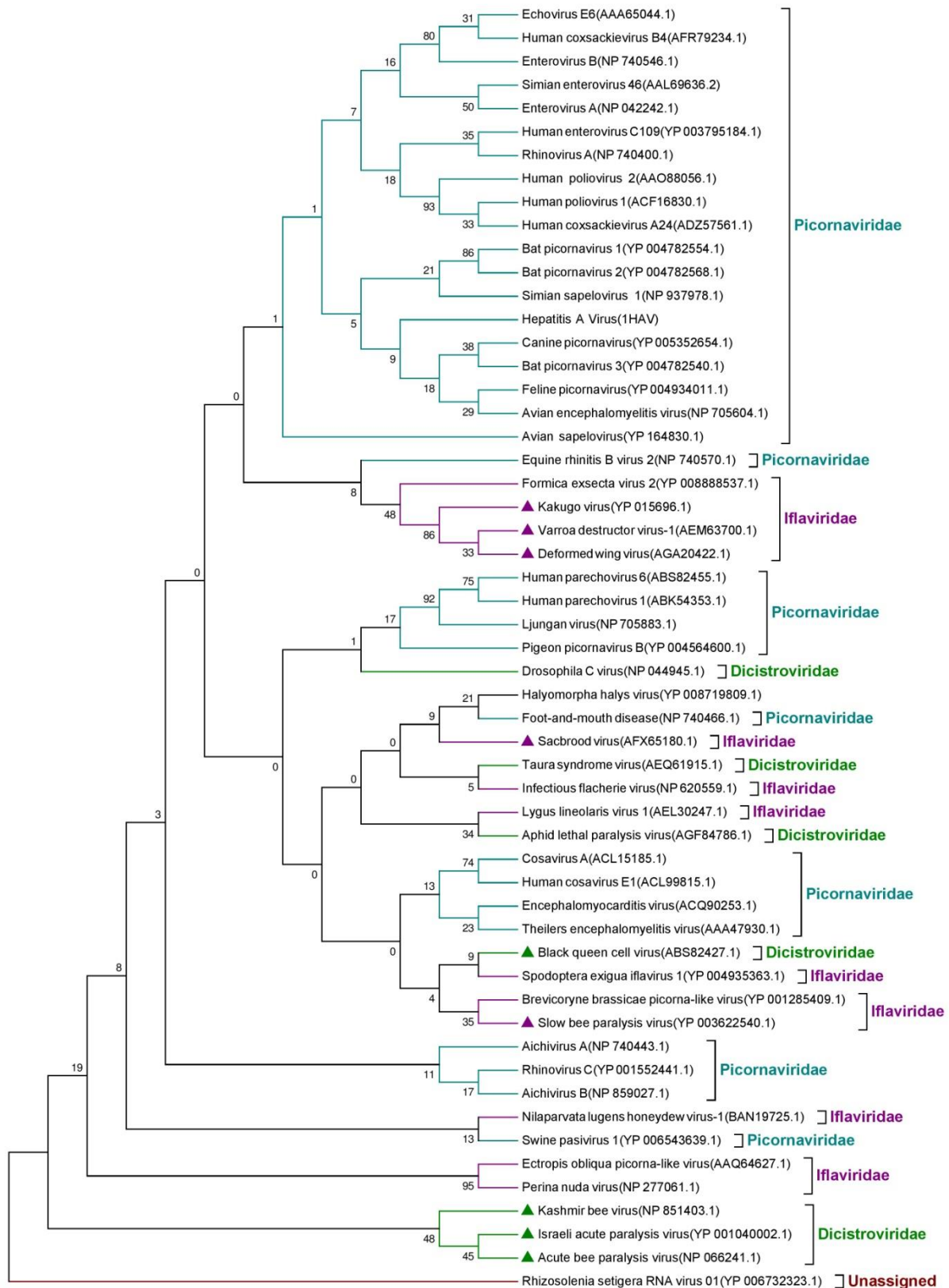


Figura 8. Árvore filogenética do motivo 4 construída pelo método de Máxima Verossimilhança.