

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

**CARACTERIZAÇÃO MOLECULAR DO VÍRUS DA ANEMIA
INFECCIOSA EQUINA NO PANTANAL E PADRONIZAÇÃO DE
qPCR PARA DIAGNÓSTICO**

CAMILA DANTAS MALOSSI

BOTUCATU - SP

Fevereiro 2019

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

**CARACTERIZAÇÃO MOLECULAR DO VÍRUS DA ANEMIA
INFECCIOSA EQUINA NO PANTANAL E PADRONIZAÇÃO DE
qPCR PARA DIAGNÓSTICO**

CAMILA DANTAS MALOSSI

Tese apresentada ao Instituto de
Biociências, Universidade Estadual
Paulista “Júlio de Mesquita Filho”,
Campus de Botucatu, para obtenção do
título de Doutora em Biotecnologia.

Orientador: Prof. Dr. João Pessoa Araujo Júnior

Coorientadora: Dra. Leila Sabrina Ullmann

BOTUCATU - SP

Fevereiro 2019

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Malossi, Camila Dantas.

Caracterização molecular do vírus da anemia infecciosa equina no pantanal e padronização de qPCR para diagnóstico / Camila Dantas Malossi. - Botucatu, 2019

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: João Pessoa Araujo Junior

Coorientador: Leila Sabrina Ullmann

Capes: 21200009

1. Anemia infecciosa equina. 2. Pantanal Mato-grossense (MS e MT). 3. Vírus da anemia infecciosa equina. 4. Diagnóstico. 5. Sequenciamento completo do genoma.

Palavras-chave: Anemia Infecciosa Equina; Pantanal; Vírus da Anemia Infecciosa Equina; diagnóstico; sequenciamento genômico.

AUTOR: Camila Dantas Malossi.

TÍTULO DO TRABALHO: Caracterização molecular do Vírus da Anemia Infecciosa Equina no Pantanal e padronização de qPCR para diagnóstico

LOCAL: IBTEC/UNESP – Campus de Botucatu/SP.

COMISSÃO EXAMINADORA

Prof. Dr. João Pessoa Araujo Júnior

Presidente e Orientador

Departamento de Microbiologia e Imunologia

Instituto de Biociências – UNESP – Botucatu – São Paulo

Prof. Dr. Ricardo Luiz Moro de Sousa

Membro Titular

Departamento de Medicina Veterinária

Faculdade de Zootecnia e Engenharia de Alimentos – USP – São Paulo

Dra. Flavia Hebler Barbosa Trovão

Membro Titular

Hemocentro de Botucatu

Faculdade de Medicina de Botucatu – UNESP – Botucatu – São Paulo

Prof. Dra. Maria Inês de Moura Campos Pardini

Membro Titular

Departamento de Clínica Médica

Faculdade de Medicina de Botucatu – UNESP – Botucatu – São Paulo

Prof. Dr. José Paes de Oliveira Filho

Membro Titular

Departamento de Clínica Veterinária

Faculdade de Medicina Veterinária e Zootecnia – UNESP – Botucatu – São Paulo

Prof. Dra. Rejane Maria Tommasini Grotto

Membro Suplente

Departamento de Bioprocessos e Biotecnologia

Faculdade de Ciências Agrônômicas – UNESP – Botucatu – São Paulo

Prof. Dr. Angelo Jose Magro

Membro Suplente

Departamento de Bioprocessos e Biotecnologia

Faculdade de Ciências Agrônômicas – UNESP – Botucatu – São Paulo

Prof. Dr. Marcos Bryan Heinemann

Membro Suplente

Departamento de Medicina Veterinária Preventiva e Saúde Animal

Faculdade de Medicina Veterinária e Zootecnia – USP – São Paulo

Botucatu, 22 de fevereiro de 2019.

Agradecimentos

Ao Prof. João Pessoa Araujo Junior, pela oportunidade, confiança, orientação e paciência. Obrigada por sempre me dar a oportunidade de dizer o que penso, e mesmo tentar um experimento que você sabe que não dará certo. Todos nós do laboratório crescemos muito no Laboratório de Diagnóstico Molecular.

Aos professores e pesquisadores que me auxiliaram na coleta de amostras e tornaram tudo isso possível: Márcia Furlan e toda a equipe da Embrapa, e prof Daniel Moura de Aguiar e toda a equipe da UFMT. Obrigada pelas discussões sobre AIE, pelos telefonemas, pelas viagens da Márcia para nos trazer as amostras. Se esse trabalho está finalizado hoje, vocês tem tanto mérito quanto eu.

À FAPESP, EMBRAPA e FUNDIBIO, pelo auxílio financeiro para a realização deste trabalho.

Falando em coleta de amostras, devo meu respeito e sinceros agradecimentos aos animais Malagueta e Luna, cujas amostras foram essenciais para esse projeto. Sei que nenhum de vocês está mais nesse plano e suas existências irão ajudar muitas outras aqui neste mundo. Obrigada!

Ao prof. Ângelo Magro, pelo auxílio nas análises de bioinformática e proteínas. Obrigada por responder todas as minhas perguntas, mesmo as mais simples!

Ao Jedson Cardoso, que se prontificou em analisar os dados do NGS em plataformas que não conhecíamos, para trazer mais resultados com a análise *denovo*. Sei que seu tempo era curto, muito obrigada!

À todos os funcionários do IBB, que sempre nos ajudaram seja recebendo amostras, seja para uma pausa gostosa para o café: Ana, Lulinha, Luiz Alquati, Rafael, Larissa, Ivana, dentre outros.

À todos os funcionários do IBTEC, a segunda casa que o laboratório ganhou com mais espaço e desafios: Edna, Raquel, Mary, Bruno, Valter, dentre outros.

Aos alunos da primeira turma da Pós Graduação de Biotecnologia, que compartilharam comigo as experiências, disciplinas, aflições e alegrias de trabalhos finalizados. Vocês são incríveis e quem fez parte dessa turma sabe o que é alunos unidos pela pós. Sucesso a todos!

À todos aqueles que me ajudaram na rotina do LDMVET, seja na bancada, seja com laudos, seja com caronas: eu não conseguiria terminar esse trabalho sem o suporte de todos! Muito, mas muito obrigada!

Compartilhar um ambiente como um laboratório é uma experiência única. Tive muita sorte em ter colegas de laboratório que tornaram essa experiência maravilhosa. A todos vocês, obrigada por me manterem firme até o fim. É maravilhoso ver que vocês não são apenas colegas, mas amigos. Curan e Duroc, obrigada por me fazerem rir e repensar muitas das minhas ações para me superar sempre. Ingrid, você é o melhor presente que o Duroc trouxe para nosso grupo (rs) e sou grata por ter conhecido alguém com sua garra, mesmo com você mesma duvidando dela. Michely, sinto falta de você todos os dias em nosso lab para xingarmos a vida, mas torço todos os dias pelo seu sucesso. Raíssa, obrigada por iniciar esse projeto com AIE e por passá-lo para mim depois de seu mestrado. Arrase na sua nova jornada! Ariani e Jacqueline, obrigada pelos momentos de aprendizado enquanto vocês trabalhavam em seus doutorados. Taís, você me ensinou bem mais do que um ELISA e um qPCR de diagnóstico, e estamos todos no laboratório rezando por sua saúde. Claudia Filoni, você foi uma mãe para todos nós em seu período aqui no LDM. Tenha certeza que cresci enormemente como pessoa com seus conselhos. Clau Tozato, obrigada por ajudar a todos sem medir esforços. Todos ficamos muito tristes e muito felizes também com sua mudança para o exterior. Desejo a você toda felicidade do mundo lá. Marianna, obrigada por sempre estar disposta a ajudar e ensinar a todos. Aprendi mais sobre pescados com você do que durante toda a minha graduação! Luana, obrigada pela sua energia no lab, estamos todos torcendo pelo seu sucesso! Leticia, foi um prazer te ajudar no doutorado; desejo-lhe que encontre a felicidade em qualquer caminho que você escolher (e que o próximo Tróia nos aguarde!). Guilherme, obrigada pela força no laboratório e nas corridas (Tróia) também! Cristiano, obrigada por todas as palavras de incentivo quando as coisas não davam certo (e pelos chocolates também!). Por favor, continue essa pessoa que nos passa tranquilidade e paz. Cid, muito obrigada por complementar o nosso laboratório! Esse laboratório cheio de mulheres precisava de alguém como você para nos ajudar a ter equilíbrio (e brejas também!).

Tenho um agradecimento especial para as meninas que formaram o “Lab da bagunça”. Pam, Jéssica, Jaqueline, Marina, Leila e Camila. Todas vocês não só me ajudaram enormemente com meu doutorado, mas me ajudaram a sobreviver a ele. Os almoços, as noites de cinema, os exercícios, os conselhos, a companhia para fazer exame de sangue, estão todos no meu

coração. Vocês são maravilhosas e sou uma pessoa melhor sabendo que vocês estão ao meu lado.

À Prof. Leila Ullmann, ou Leilinha, que além de ser minha coorientadora e me ajudar a não desistir em nenhuma etapa do trabalho, é uma amiga valiosa que Botucatu me trouxe.

À minha vó Maria Cecília, que tem o dom de amar e enlouquecer todos os filhos e netos (rs). Obrigada pelo orgulho que você sente de mim, vó, bem como todo apoio que sempre deu.

A Daniel e Marlene, que me adotaram como filha, e me apóiam com sorrisos e boa comida para eu continuar minha jornada. Muito obrigada!!!

À pessoa que possibilitou tudo isso desde o início. Mãe, obrigada por me fazer o que sou hoje. Tenho desde pequena o melhor exemplo de luta e superação dentro de minha casa. A você devo tudo e mais um pouco. Esse doutorado é uma conquista, não só minha mas sua também. Te amo!

Junto comigo, minha mãe criou alguém que seria minha cara metade. Mana, você cresceu me fazendo crescer junto com você. Eu nunca duvidei da sua capacidade assim como você nunca duvidou da minha. Somos a melhor face uma da outra. Te amo sempre, doutora.

Por último, mas jamais menos importante, obrigada por tudo, Luca. Seu apoio não se resume a não me deixar desistir e suportar minha ausência e mau humor. Esta tese está escrita graças a você. Você ajudou desde a repor ponteiros até emitir laudos. Obrigada por deixar meu mundo mais colorido e melhor para vivê-lo.

**“I close my eyes and I can see
A world that's waiting up for me
That I call my own
Through the dark, through the door
Through where no one's been before
But it feels like home
They can say, they can say it all sounds crazy
They can say, they can say I've lost my mind
I don't care, I don't care, so call me crazy
We can live in a world that we design
'Cause every night I lie in bed
The brightest colors fill my head
A million dreams are keeping me awake
I think of what the world could be
A vision of the one I see
A million dreams is all it's gonna take
Oh a million dreams for the world we're gonna make [...]”**
A million dreams - The Greatest Showman (2018)

Resumo

MALOSSI, C. D. Caracterização molecular do Vírus da Anemia Infecciosa Equina no Pantanal e padronização de qPCR para diagnóstico. 2019. 139 f. Tese (Doutorado) – Instituto de Biociências de Botucatu, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Botucatu, São Paulo.

O Vírus da Anemia Infecciosa Equina (VAIE) é um lentivírus que infecta equídeos causando a Anemia Infecciosa Equina. A doença possui distribuição mundial e o principal método de controle empregado é a eutanásia dos animais positivos, uma vez que não existe cura ou vacina disponível. No Brasil, a região do Pantanal é endêmica para a doença e, somente nessas áreas, a eutanásia não é obrigatória. A alta variação genética de lentivírus durante uma infecção prolongada é bem conhecida em várias espécies e também descrita em sequências de VAIE de outros países. Não há sequência genômica brasileira completa descrita ou mesmo um estudo da variação genética do vírus em um ambiente endêmico como o Pantanal Brasileiro. Com isso, este trabalho teve como objetivos a caracterização molecular do VAIE em equinos da região do Pantanal Brasileiro, e a padronização da PCR quantitativa (qPCR) para detecção do DNA proviral de VAIE. Duas amostras de plasma equino positivas para VAIE foram submetidas ao sequenciamento do genoma completo utilizando uma combinação de técnicas tradicionais e de nova geração. O genoma a partir do RNA viral foi obtido, anotado e comparado com sequências virais obtidas de animais naturalmente infectados em outros países. As sequências provenientes do Pantanal apresentaram identidade entre 75 e 77 % com sequências de outros países e 88,54 % entre si, classificando-as em um clado individual na árvore filogenética dos vírus já descritos, exibindo a maior variação genética descrita entre sequências de um mesmo país. Com a análise do gene do envelope viral, observou-se que os dois animais estudados estavam infectados por diferentes quasiespécies virais. Baseado nas sequências obtidas, um qPCR foi padronizado para amplificar 71pb do gene *tat* de VAIE. A técnica foi testada em 255 amostras de DNA total de equinos do Pantanal e comparada com os resultados obtidos em testes sorológicos (IDGA e ELISA) e uma Semi Nested PCR. A qPCR apresentou melhor sensibilidade, especificidade e alta correlação com os resultados da Semi Nested PCR, sendo ainda um teste com menor risco de contaminação e melhor praticidade. A qPCR detectou o mesmo número de amostras positivas

que o padrão ouro IDGA, porém ambos detectaram amostras diferentes, confirmadas por sequenciamento. Um fluxograma diagnóstico foi proposto para diminuir o risco de falsos negativos em áreas endêmicas, com o uso de IDGA, seguido de ELISA e qPCR.

Palavras-chave: Vírus da Anemia Infecciosa Equina, Anemia Infecciosa Equina, Pantanal, área endêmica, sequenciamento genômico, diagnóstico, equinos.

Abstract

MALOSSO, C. D. Molecular characterization of Equine Infectious Anemia Virus from Pantanal and qPCR standardization for diagnosis. 2019. 139 p. Thesis (Doctorate) – Botucatu Biosciences Institute, São Paulo State University “Júlio de Mesquita Filho”. Botucatu, São Paulo.

Equine infectious anemia virus (EIAV) is a lentivirus that infects equids and causes equine infectious anemia (EIA). This disease presents worldwide distribution and the main control method is the euthanasia of positive animals, since there is no cure or vaccine available. The Pantanal region in Brazil is endemic for the disease, and euthanasia is not mandatory in this area. The high lentivirus genetic variation during a prolonged infection is well known in several species and also described in EIAV sequences in other countries. There are no Brazilian viral genomic sequences available or even a study of genetic variation in an environment such as Brazilian Pantanal. This work aimed to characterize the EIAV complete genome in equines from Brazilian Pantanal and to standardize a quantitative real time PCR (qPCR) for detection of EIAV proviral DNA. Two plasma samples of EIAV positive horses were submitted to massive sequencing by combination of Sanger and NGS sequencing. The complete genome from viral RNA was obtained, annotated and compared with viral sequences of naturally infected animals from other countries. Sequences from Pantanal showed identity between 75% and 77% with other countries sequences and 88.54% with each other, classifying them into an individual cluster in EIAV cladogram and exhibiting the greatest genetic variation between sequences from the same country. With viral envelope gene analysis, both Brazilian sequences were classified as two quasispecies. A qPCR based on the sequences obtained was standardized to amplify 71 bp of *tat* gene of EIAV. This technique was tested in 255 equine total DNA from Pantanal and compared with results of serological tests (AGID and ELISA) and a Semi Nested PCR. Quantitative PCR presented better sensitivity, specificity and high correlation with Semi Nested PCR results, still considered a test with lower risk of contamination and better practicality. The qPCR detected the same number of positive samples as the gold standard AGID, but both detected different samples, confirmed by sequencing. A diagnosis flowchart was proposed to decrease false-negative risks in endemic areas, using a sequence of AGID, ELISA and qPCR.

Key words: Equine Infectious Anemia Virus, Equine Infectious Anemia, Pantanal, endemic área, genome sequencing, diagnosis, equine.

Sumário

1	Revisão bibliográfica	14
1.1	Histórico	14
1.2	Agente etiológico: o vírus da Anemia Infecciosa Equina	16
1.2.1	Genoma de VAIE: poliproteína gag	17
1.2.2	Genoma de VAIE: poliproteína pol	19
1.2.3	Genoma de VAIE: poliproteína env	22
1.2.4	Genoma de VAIE: genes acessórios	23
1.2.5	Genoma de VAIE: long terminal repeats (LTR)	25
1.3	Ciclo de replicação do vírus da Anemia Infecciosa Equina	26
1.4	Patogenia e clínica	33
1.5	A transmissão da anemia infecciosa equina	35
1.6	Diagnóstico	37
1.7	Evolução viral	39
1.8	Genomas de VAIE sequenciados	41
1.9	VAIE no Pantanal, Brasil	42
2	Justificativa	46
3	Objetivos	48
4	Artigo Científico 1	49
	High variations in <i>Equine infectious anemia virus</i> genome from naturally infected horses in Pantanal, Brazil: an endemic region case	49
5	Artigo Científico 2	73
	A quantitative PCR to <i>Equine infectious anemia virus</i> from Brazil: evidences of serological silent equids require a new diagnosis workflow to EIA control	73
6	Conclusões finais	93
7	Referências bibliográficas	94
8	Anexos	105
9	Apêndices	107

1 Revisão bibliográfica

O vírus da Anemia Infecciosa Equina (VAIE) é o agente etiológico da Anemia Infecciosa Equina (AIE), uma doença considerada um desafio para a medicina veterinária desde sua descoberta. Afetando equídeos em geral e sendo uma doença persistente por toda a vida do animal (como observado em todas as doenças acometidas por lentivírus), os estudos até hoje realizados ainda não chegaram à cura ou vacina eficaz. Com isso, um diagnóstico com alta acurácia é essencial para o controle da doença.

1.1 Histórico

O primeiro relato da doença data de 1843 na França (LIGNÉ, 1843), e em 1904 foi relacionada com um agente filtrável que persistiam no hospedeiro equídeo e era transmitida pelo sangue (VALEE; CARRE, 1904). Portanto, o vírus da anemia infecciosa equina (VAIE) foi o primeiro retrovírus animal a ser atribuído como agente etiológico de uma enfermidade viral, precedendo em alguns anos as principais descobertas de agentes filtráveis, como a leucemia aviária (ELLERMAN; BANG, 1908) e o sarcoma aviário (ROUS, 1910). A transmissão mecânica de VAIE por insetos hematófagos vetores (especificamente tabanídeos) foi demonstrada em 1942 (STEIN et al., 1942) e em 1967, o vírus foi pela primeira vez propagado em cultura de leucócitos de equino (KOBAYASHI; KONO, 1967). A limitação imposta pela falta do cultivo do vírus restringiu muito o conhecimento a respeito do vírus e consequentemente da doença (LEROUX; CADORÉ; MONTELARO, 2004). Dessa forma, práticas de controle para AIE somente foram possíveis a partir de 1972 quando foi padronizado o teste de imunodifusão em gel de ágar (IDGA) que detecta anticorpos contra o vírus usando antígeno extraído de baço de equinos infectados (COGGINS et al., 1972).

Mais de 100 anos depois da descoberta da doença, em 1952, Manete admitiu a ocorrência da doença no Brasil (REIS; LEITE, 2016). Em 1967, no Jockey Club do Rio de Janeiro, Dupont relatou o primeiro caso no país em exame *post mortem* de um equino (REIS; LEITE, 2016) e a doença também foi constatada no Rio Grande do Sul (GUERRERO et al.,

1968). Em 1984, Montagnier observou que o soro de equinos positivos para AIE reconhecia proteínas virais em soro de pacientes com HIV e descreveu a relação antigênica entre as doenças (REIS; LEITE, 2016).

No Brasil, o diagnóstico, as ações de profilaxia e controle, a erradicação dos focos, o controle de trânsito dos animais e outras normas estão pautados na Instrução Normativa 45 (IN45) de 15 de junho de 2004, do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) (BRASIL, 2004). Nela, está determinada a notificação obrigatória, marcação e eutanásia de animais positivos confirmados pelo teste de IDGA em laboratórios credenciados. As propriedades identificadas com animal(is) positivo(s) devem ser interditadas e o isolamento dos demais deve ser realizado, bem como sorologia do efetivo equídeo restante do local. A contraprova é facultativa e não há indenização.

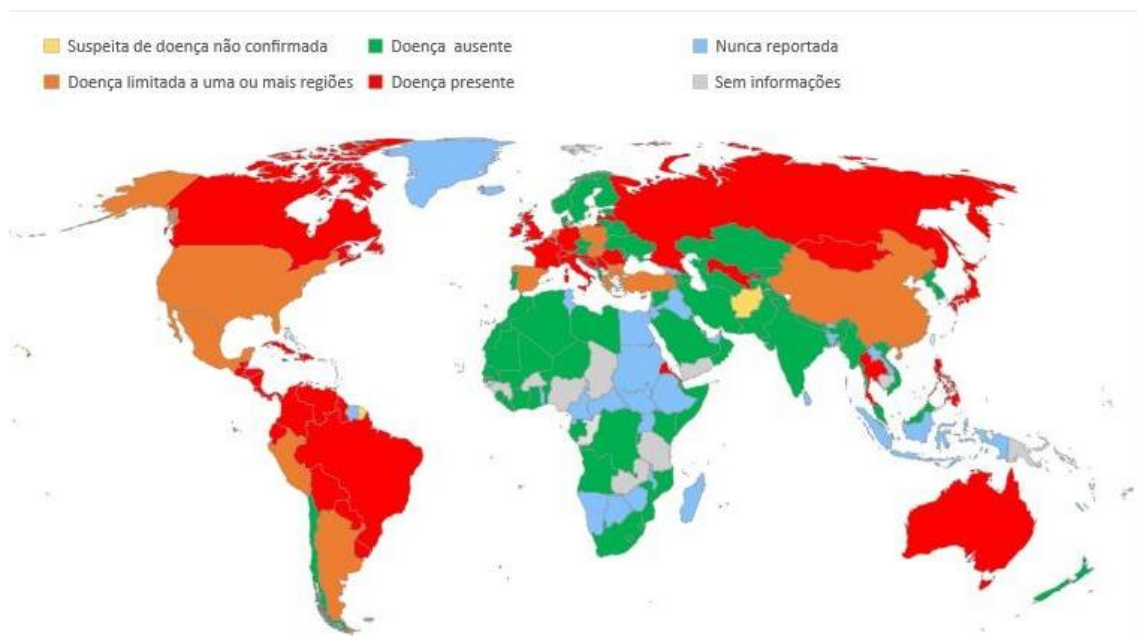


Figura 1: Situação da Anemia Infecciosa Equina reportada mundialmente de 2005 a 2018 segundo o Banco de Dados de Informação sobre Saúde Animal Mundial (*World Animal Health Information Database - WAHIS Interface*) da Organização Mundial de Saúde (OIE) - Disponível em:

<http://www.oie.int/wahis_2/public/wahid.php/Countryinformation/Countrytimelines>.

Acesso em: 08/10/2018.

Atualmente, equídeos infectados com anemia infecciosa equina são encontrados no mundo todo e reportados frequentemente em países da Europa, Ásia, Oceania e Américas. A OIE não apresenta relatos recentes de AIE na África, exceto um caso em março de 2013 em Shambiko, Eritrea (OIE, 2018). Desde 2005, 55 países reportaram a presença da doença no mundo (Figura 1) (OIE, 2018).

1.2 Agente etiológico: o vírus da Anemia Infecciosa Equina

O vírus da Anemia Infecciosa Equina é classificado como membro do gênero *Lentivirus*, da subfamília *Orthoretrovirinae*, família *Retroviridae* e ordem *Ortervirales* sendo por isso denominado muitas vezes como “primo-irmão” do vírus da Imunodeficiência Humana (HIV) (LEROUX; CADORÉ; MONTELARO, 2004). Dentro do mesmo gênero, podemos citar além do HIV1 e HIV2, o vírus da Imunodeficiência Bovina (BIV), o vírus da doença de Jembrana, o vírus da Artrite-Encefalite Caprina (CAEV), o vírus da Imunodeficiência Felina (FIV), o lentivírus de Puma, o vírus da Imunodeficiência Símia (SIV) e o vírus Maedi-Visna (ICTV, 2018).

A análise do vírus da Anemia Infecciosa Equina por microscopia eletrônica revela uma partícula oval ou circular de aproximadamente 115 nm de diâmetro (MATHEKA et al., 1976; WEILAND et al., 1977). É um vírus envelopado com proteínas de matriz recobrindo externamente o nucleocapsídeo icosaédrico, mediando a associação com a superfície interna do envelope (COOK; LEROUX; ISSEL, 2013). Cada vírion possui duas cópias do genoma de fita simples de RNA com sentido positivo. Com 8,4 kb, o genoma de VAIE é um dos menores dentre os lentivírus (ISSEL et al., 2014). O DNA proviral consiste de três principais genes estruturais de retrovírus (*gag*, *env* e *pol*) com longas terminações repetitivas (LTR) nas pontas 5' e 3' da sequência gênica. Por ser um retrovírus classificado como complexo (assim como todo lentivírus), VAIE possui outras ORFs além das três principais, codificando proteínas acessórias com funções regulatórias, de escape do sistema imune do hospedeiro ou mesmo de aumento da patogenicidade (ISSEL et al., 2014). Comparado a outros vírus do mesmo gênero o VAIE possui o menor número de ORFs acessórias (três contra quatro em BIV e seis em HIV). Duas delas são comuns a todos os lentivírus (*tat* e *rev*), enquanto uma (*S2*) é exclusiva

em VAIE (ISSEL et al., 2014). O VAIE é o único membro dos lentivírus que não possui uma ORF que codifique um ortólogo para *vif*, uma proteína dirigida contra uma defesa do hospedeiro para retrovírus, da família do complexo 3 da apolipoproteína β editante (APO β EC3). A ausência dessa proteína é inesperada, já que equinos possuem mais genes de APO β EC3 que qualquer outra espécie não primata (BOGERD et al., 2008).

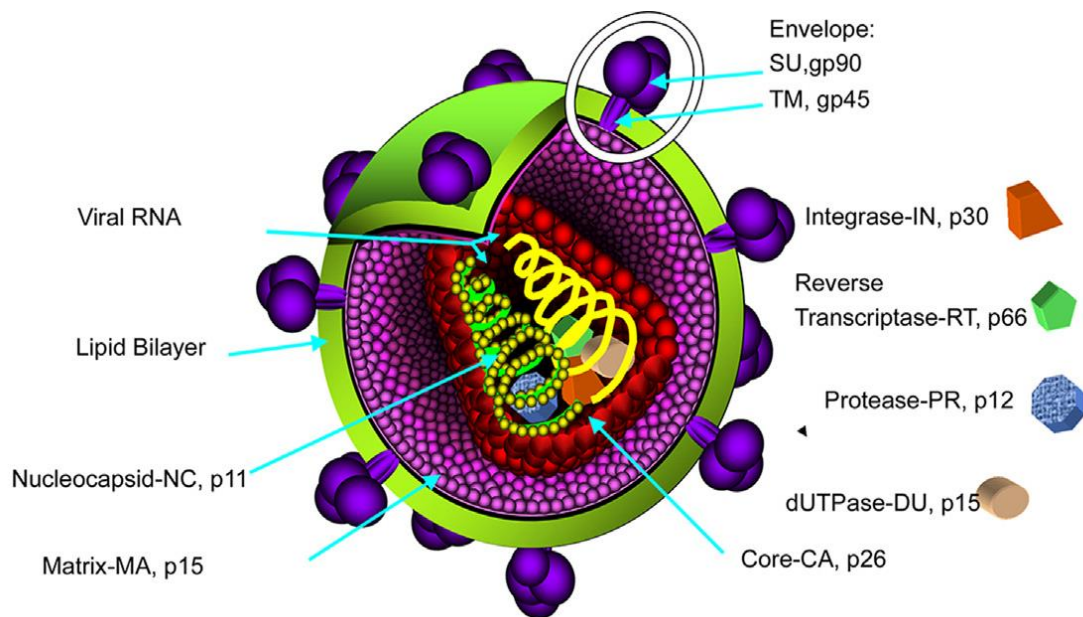


Figura 2: Estrutura do vírus da Anemia Infecciosa Equina mostrando as proteínas estruturais, RNA genômico e envelope. Fonte: ISSEL et al. (2014).

1.2.1 Genoma de VAIE: poliproteína gag

O gene *gag* de VAIE codifica a matriz (MA – p15), o capsídeo ou core (CA – p26), o nucleocapsídeo (NC – p11) e a proteína p9. Estas proteínas são responsáveis pela montagem e liberação das partículas virais da célula hospedeira. Caso expressa sozinha, a poliproteína *gag* pode iniciar a montagem e o brotamento de partículas semelhantes a vírus (*virus like particles* - VLPs) não infecciosas da membrana celular (HATANAKA et al., 2002). A poliproteína *gag* é traduzida a partir do RNAm genômico, sem alterações de frame ou *splicing*, e posteriormente é clivada pela protease viral (COOK; LEROUX; ISSEL, 2013).

A proteína de matriz MA liga-se na membrana lipídica do hospedeiro e estimula a montagem do vírus (na sua forma *gag* não processada). A agregação dos precursores *gag* na

membrana induz a projeção desta e vírus podem realizar o brotamento (FIELDS; KNIPE; HOWLEY, 2007). No vírion maduro, o MA liga o envelope ao capsídeo. A estrutura cristalográfica do MA foi determinada e mostrou que apesar de uma sequência primária bem diferente dos MA de SIV e HIV, as estruturas são similares, com apenas as regiões N e C terminal mais próximas (HATANAKA et al., 2002). O MA de HIV desempenha um papel na montagem de partículas virais por ligação a fosfatidilinositol- 4,5-bifosfato (PIP2) localizado no interior da membrana plasmática da célula hospedeira, uma associação que é mediada pelo miristato N-terminal. Apesar da ausência desta modificação N-terminal, o MA de VAIE também se prende à superfície interna das membranas, mas estudos conduzidos *in vitro* sugerem que pode se ligar com maior afinidade ao fosfatidilinositol 3-fosfato (PIP3) do que o PIP2. O MA também contém um sinal de exportação nuclear rico em leucina semelhante ao HIV-1: este sinal é possivelmente necessário para indicar a localização nuclear do MA e assegurar a disponibilidade citoplasmática (HATANAKA et al., 2002).

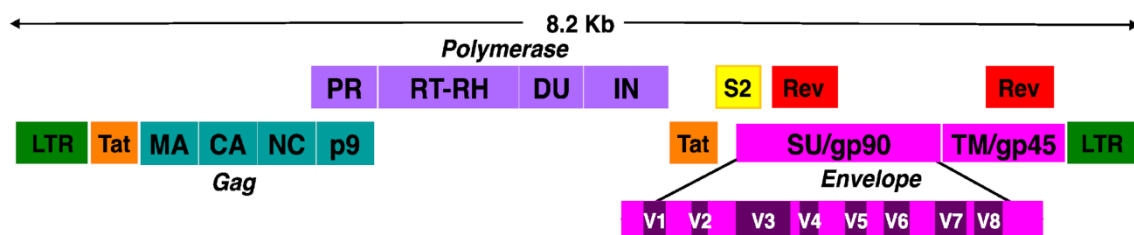


Figura 3: Esquema do genoma proviral do vírus da Anemia Infecciosa Equina mostrando os genes (itálico) e as respectivas proteínas virais codificadas (negrito em preto). As regiões hipervariáveis V1 a V8 da proteína de superfície (SU/gp90) estão indicadas em branco. Fonte: CRAIGO; MONTELARO (2013).

A proteína de capsídeo CA é o componente que forma um invólucro protéico em torno do genoma e desempenha um papel na remoção do revestimento, montagem do vírus e brotamento (GRUND et al., 1994). O CA apresenta uma região de similaridade com outros lentivírus, conhecido como “região de homologia principal” ou *major homology region* (MHR) em inglês (GRUND et al., 1994). Também apresenta epítomos para linfócitos T citotóxicos, apresentando alta imunogenicidade (CHUNG; MEALEY; MCGUIRE, 2005).

A proteína do nucleocapsídeo NC se liga ao RNA genômico formando um complexo que está envolvido no brotamento do vírus, importação e integração nuclear (JIN et al., 2005; LEROUX; CADORÉ; MONTELARO, 2004; PAYNE; FULLER, 2010). A sequência de NC contém dois domínios ligantes de zinco do tipo CCHC (CX₂CX₄HX₄C) separados por uma curta sequência de ligação que permite ao NC ligar-se ao genoma viral via sinal de empacotamento (ϕ) (AMODEO et al., 2006). A proteína NC também demonstrou possuir atividade de topoisomerase, sugerindo um papel na integração e transcrição de DNA após a entrada viral (MATSRAFI et al., 1996).

A última proteína codificada por *gag* é a proteína p9, chamada assim por apresentar 9 kDa, com correspondência a p6 codificada pelo *gag* de HIV (JIN et al., 2005). Apesar de ainda não possuir função conhecida, sabe-se que a p9 desempenha um papel importante na produção proviral e infecção precoce (JIN et al., 2005). A proteína p9 é variável dentre as sequências já descritas de VAIE, no entanto o único domínio encontrado na proteína, chamado YPDL tardio, é conservado em todos os isolados virais, sugerindo sua importância na sequência (PUFFER et al., 1997; PUFFER; WATKINS; MONTELARO, 1998).

1.2.2 Genoma de VAIE: poliproteína pol

O gene *pol* codifica a protease (PR – p12), a transcriptase reversa com atividade de RNase H (RT-RH – p66), a dUTPase (DU – p15) e a integrase (IN – p30). Assim como *gag*, *pol* é codificado no RNAm genômico, porém em um quadro de leitura diferente de *gag*. Um *frame shifting* ribossomal permite a mudança de base em -1, de modo a traduzir *pol* após a tradução de *gag* (COOK; LEROUX; ISSEL, 2013).

Uma vez que enzimas tendem a ser conservadas e os genes de *pol* em lentivírus compartilham uma homologia de sequência relativamente alta, a identificação de motivos de VAIE é geralmente extrapolada de trabalhos com HIV e SIV (COOK; LEROUX; ISSEL, 2013). As sequências de *pol* em VAIE conhecidas hoje derivam, em sua maioria, das sequências genômicas completas obtidas para o vírus (COOK; LEROUX; ISSEL, 2013). Dessa forma, a literatura possui menor quantidade de sequências descritas para *pol* do que para *gag* e *env*, que são mais estudadas por sua função no diagnóstico. Essas moléculas possuem estruturas similares, embora não sequências primárias de nucleotídeos. Muitas

dessas alterações de nucleotídeos não resultam em um diferente aminoácido e, quando este é alterado, não modifica muito a estrutura terciária, quando comparado a outros retrovírus.

As proteases retrovirais são enzimas chave na propagação viral e são inicialmente sintetizadas em uma poliproteína subsequentemente clivada pela atividade da protease viral pré-existente no vírion da infecção inicial (DUNN et al., 2002). A protease PR também é responsável pela clivagem de *gag* e *pol* para formar o vírion maduro. Estas proteínas são alvo de muitas drogas anti HIV e algumas mutações são observadas em linhagens virais resistentes a estas drogas (TÖZSÉR, 2010). As proteases retrovirais ativas são homodímeros, com cada dímero estruturalmente relacionado à classe maior de peptidases aspárticas de cadeia simples. Cada monômero tem quatro elementos estruturais: duas alças distintas, uma alça larga contendo o ácido aspártico catalítico e uma hélice D (DUNN et al., 2002). A análise por cristalografia de raios X realizada na linhagem EIAV_{WY} demonstrou que a única grande diferença entre essa poliproteína e outras proteínas retrovirais é a presença de uma segunda alfa-hélice dentro da região C terminal da molécula (GUSTCHINA et al., 1996). Dentre as linhagens de VAIE, a sequência de aminoácidos é muito conservada, exceto por uma substituição do aminoácido 54, que é leucina (L) em EIAV_{MIY}, isoleucina (I) em EIAV_{WY} e valina (V) em EIAV_{LIA} e EIAV_{IRE} (DONG; COOK; ZHU, 2014).

A transcriptase reversa de VAIE compartilha diversos motivos importantes com a mesma enzima do HIV e uma alta conservação de aminoácidos, particularmente aqueles com uma carga positiva ou negativa (DONG; COOK; ZHU, 2014). Dois resíduos de aspartato potencialmente catalíticos consecutivos (RUNGROTMONGKOL; MULHOLLAND; HANNONGBUA, 2007) nas posições 185 e 186 estão também presentes na RT de VAIE. No entanto, o motivo DVGD em HIV-1, que também pode desempenhar um papel importante na reação de polimerização de nucleotídeos, é substituído por DIGD em VAIE (DONG; COOK; ZHU, 2014).

A extremidade 3' da RT de VAIE codifica um domínio de RNase H (RH) que catalisa a degradação de RNA quando hibridizado com DNA (RNA:DNA), produzido durante a transcrição reversa (FIELDS; KNIPE; HOWLEY, 2007). O HIV-1 RH contém um motivo rico em glicina (AWVPAHKGIGGN), que é altamente conservado entre os lentivírus. As únicas alterações em VAIE são a segunda alanina e a segunda glicina alterada para uma glicina e tirosina, respectivamente (DONG; COOK; ZHU, 2014). Além de um motivo rico em

glicina (posições de aminoácidos 534-545), os quatro resíduos ácidos (D443, E478, D498, D549) que são altamente conservados entre os diferentes retrovírus (DAVIES et al., 1991) e considerados essenciais para a catálise enzimática e a ligação de íons metálicos na RNase H são presentes nas linhagens de VAIE obtidas a campo (DONG; COOK; ZHU, 2014).

Lentivírus replicam-se em células não proliferativas, fato que os fez, incluindo o VAIE, atraente como potenciais vetores de terapia gênica (LIU et al., 2015). Como as células não proliferativas não sofrem extensa síntese de DNA, elas geralmente possuem baixos níveis de desoxinucleotídeos intracelulares (dNTP), com exceção do trifosfato de desoxiuridina (dUTP) que se acumula devido à expressão limitada da hidrolase de dUTP (dUTPase). A combinação de concentrações baixas de dNTP intracelular mais níveis elevados de dUTP em células não proliferativas constitui um ambiente hostil para a replicação viral. Portanto, a dUTPase codificada por vários lentivírus, incluindo o VAIE, pode impedir a incorporação de U no genoma viral (COOK; LEROUX; ISSEL, 2013). A dUTPase (DU) de VAIE apresenta um “anel de uracila” composto por N67, G69, Q80 e I82. A ligação da uracila induz uma alteração conformacional, que também coloca o G68 em contato com a uracila. Uma outra região de ligação interage com o grupo pentose da uracila através de A17, I71, N72 e Y75. O grupo fosfato interage com D26, S58, K56, E95 e K98. A ligação ao íon metálico envolve possivelmente um grupo de resíduos, alguns dos quais estão também envolvidos na ligação do fosfato: um aglomerado de R13, D20, K56, E95 e um aglomerado carregado negativamente de E15, D16, D72 e E73 (DAUTER et al., 1999).

A integrase (IN) é a enzima que promove a integração do DNA proviral no genoma do hospedeiro, formando, junto com outras proteínas virais e celulares, o complexo de pré-integração (*preintegration complex* - PIC) (LIU et al., 2015). Como resultado, o retrovírus estabelece uma infecção latente na célula e se mantém durante todo o tempo de vida do hospedeiro (LIU et al., 2015). A IN de VAIE contém um motivo ligante de zinco HHCC conservado em lentivírus (H12, H16, C40, C43), resíduos hidrofóbicos conservados (I5, A8, L22, A33 e I36) e resíduos do sítio catalítico (D64, D116 e E152) (COOK; LEROUX; ISSEL, 2013; DONG; COOK; ZHU, 2014; QUINLIVAN et al., 2013). Por ser um vírus estudado como vetor para terapia gênica, a IN de VAIE teve 477 sítios de integração no genoma hospedeiro mapeados e se integra em genes e regiões ricas em AT, enquanto evita a

integração em locais de início de transcrição (*transcription start sites* - TSS) (LIU et al., 2015).

1.2.3 Genoma de VAIE: poliproteína env

O gene *env* codifica a glicoproteína de superfície gp90 (SU) e a glicoproteína transmembrana gp45 (TM). Elas são traduzidas como uma poliproteína e clivadas por proteases da célula hospedeira (COOK; LEROUX; ISSEL, 2013). A SU gp90 possui aproximadamente 452 aminoácidos, enquanto a TM gp45 possui aproximadamente 415 (DONG; COOK; ZHU, 2014). Ainda não há estrutura terciária completa para tais glicoproteínas apesar de serem muito estudadas por razão de diagnóstico; normalmente são comparadas com as estruturas de gp120 (SU) e gp42 (TM) de HIV. Assim como as glicoproteínas do vírus da imunodeficiência humana, gp90 é altamente glicosilada com 17 potenciais sítios de glicosilação ligados a N (NX[S/T]); gp45 possui 5 potenciais sítios de glicosilação (LEROUX; ISSEL; MONTELARO, 1997). Juntamente com a proteína de capsídeo CA, gp90 e gp45 apresentam uma alta resposta do sistema imune do hospedeiro (CRAIGO; MONTELARO, 2010).

A TM e a SU são altamente variáveis nos animais infectados, com mutações na gp90 sendo uma das principais responsáveis pela geração de mutantes de escape do sistema imune (QUINLIVAN et al., 2013). A TM gp90 apresenta oito regiões hipervariáveis (V1 - V8), classificadas por serem regiões com mais de 30 % de divergência com a sequência original analisada EIAV_{PV} (LEROUX; ISSEL; MONTELARO, 1997). Na região V3 há a predição de um *loop* estrutural semelhante ao encontrado na gp120 de HIV (BALL et al., 1992). Como esta região contém dois dos três epítomos neutralizantes reconhecidos por anticorpos monoclonais de camundongo, foi denominado como domínio neutralizante principal ou *principal neutralizing domain* - PNDs (BALL et al., 1992). Também se observa resíduos de cisteína altamente conservados em gp90 entre as linhagens, fato esperado pois as pontes dissulfeto resultantes são essenciais para o dobramento correto da proteína e sua ligação ao receptor (DONG; COOK; ZHU, 2014).

A glicoproteína TM, assim como SU, contém cisteínas e sítios de glicosilação, mas em menor número, por ser uma proteína mais protegida e inserida no vírion. A TM contém duas

regiões hidrofóbicas: a região proximal do terminal N, que pode desempenhar um papel na fusão com a membrana durante a entrada na célula, e na região transmembrana, que ancora o complexo de glicoproteína ao vírion (RICE et al., 1990).

1.2.4 Genoma de VAIE: genes acessórios

Os transcritos acessórios de VAIE *tat* (S1), S2 e *rev* (S3) são lidos em *frameshift* e/ou *splicing* alternativo (COOK; LEROUX; ISSEL, 2013).

A proteína *tat* (proteína transativadora de transcrição) de VAIE é traduzida a partir de um RNAm que sofreu *splicing* alternativo (COOK; LEROUX; ISSEL, 2013). Esta proteína auxilia na transcrição do RNAm viral pela RNA polimerase II celular a partir do DNA proviral. Para esse processo é necessária a interação de *tat* com o RNAm viral nascente formando uma estrutura específica de alça denominada elemento TAR (*trans-activation response*) (COOK; LEROUX; ISSEL, 2013). As interações entre *tat* e TAR levam ao recrutamento de fatores celulares, incluindo o fator de extensão transcricional positivo b (P-TEFb), um complexo protéico composto de CDK9 (ciclina dependente quinase 9) e ciclina T1 que aumenta a processividade pela fosforilação do domínio carboxila da maior subunidade da RNA Polimerase II (SUÑÉ et al., 2000). O *tat* de VAIE contribui para a restrição vírus-célula hospedeira porque suas interações são limitadas aos ortólogos equinos e caninos da ciclina T (COOK; LEROUX; ISSEL, 2013). Assim como observado em outros lentivírus, o gene *tat* inicia-se com um códon de iniciação CTG. Como tal códon também inicia a tradução da proteína *rev*, este códon é importante para o posterior *splicing* alternativo em ambos transcritos. Modificar o códon de iniciação para ATG leva a uma grande redução na tradução de *rev* (CARROLL; DERSE, 1993). A proteína *tat* de VAIE consiste em uma região amino terminal, uma região central, uma região básica e o carboxi terminal, mas não possui uma região rica em cisteína encontrada em outros lentivírus (CARROLL; MARTARANO; DERSE, 1991). Esta molécula parece ser relativamente conservada (80% de identidade de sequência de aminoácidos prevista) entre as linhagens, particularmente nos domínios amino-terminais e regiões centrais (DONG; COOK; ZHU, 2014).

A proteína acessória *rev* é responsável por facilitar a exportação nuclear de RNAm viral com *splicing* incompleto e, quando ausente, apenas o RNAm bicistrônico *tat/rev* está

presente no citoplasma (COOK; LEROUX; ISSEL, 2013). Esta proteína é traduzida dos exons 3 e 4 do RNAm bicistrônico tat/rev através do códon CTG (CARROLL; DERSE, 1993). Toda a sequência responsável pela codificação de *rev* está dentro do gene de *env*, com o exon 3 no mesmo quadro de leitura de gp90, enquanto o exon 4 difere do quadro de leitura de gp45 (Figura 3). A função facilitadora de *rev* se dá pela ligação ao RNA viral através do elemento de resposta a *rev* (*rev-response element* - RRE) e então sofre multimerização para a passagem através da membrana nuclear via nucleoporina usando um sinal de exportação nuclear (*nuclear export signal* - NES) (LEE et al., 2008). O *rev* do VAIE e do FIV utilizam um NES atípico que difere de outras proteínas *rev* pela falta de um tetrâmero central e a organização de resíduos hidrofóbicos (HARRIS et al., 1998). Outra diferença é a presença de dois RRE em VAIE, RRE1 e RRE2, em vez de um RRE único com formato de alça estruturada, como visto em HIV. O RRE1 consiste de 555 nucleotídeos abrangendo o exon 3 no mRNA bicistrônico tat/rev, enquanto o RRE2 ocorre *downstream* e tem 1698 nucleotídeos de comprimento (COOK; LEROUX; ISSEL, 2013). O RRE1 contém o potenciador de *splicing* de exon (*exon splicing enhancer* - ESE) rico em purinas com 57 nucleotídeos que pode ligar o *rev* de VAIE e mediar de forma limitada a exportação nuclear de RNA incompletamente processado (BELSHAN et al., 2001). A comparação das sequências de aminoácidos de *rev* revela identidades de 68% a 73% entre as linhagens EIAV_{WY}, EIAV_{IRE}, EIAV_{LIA} e EIAV_{MIY}, e apesar das sequências dentro do domínio N-terminal variarem entre estas linhagens, observamos resíduos hidrofóbicos conservados (L36, W41, L45, L49) necessários para a função de exportação nuclear de *rev* (BELSHAN et al., 2001; HARRIS et al., 1998). Já nos resíduos de aminoácidos C-terminais implicados na ligação do RNA existem variações entre as linhagens (HARRIS et al., 1998; LEE et al., 2006) sugerindo que os vírus associados aos grupos monofiléticos separados podem apresentar diferentes sequências primárias de aminoácidos para lidar com os mesmos problemas biológicos associados replicação e persistência nos hospedeiros equídeos (COOK; LEROUX; ISSEL, 2013).

A proteína S2 é exclusiva do VAIE: possui 65-68 aminoácidos e aproximadamente 7 kDa (COVALEDA et al., 2010). Embora sua função não esteja totalmente determinada, estudos de mutagênese demonstraram que não é essencial para a replicação *in vitro* (LI; PUFFER; MONTELARO, 1998), mas é um importante determinante da virulência *in vivo* (FAGERNESS et al., 2006; LI et al., 2000). Na ausência de S2, as cargas virais associadas ao plasma são mais baixas do que aquelas observadas em animais inoculados com vírus do tipo

selvagem, indicando alteração na virulência (LI et al., 2000). Este fenótipo atenuado resultou em variantes de VAIE com deleção de S2 sendo avaliadas como vacinas vivas modificadas (LI et al., 2003), já que verificou-se que a linhagem da vacina atenuada de VAIE com múltiplas mutações no gene S2 não irá facilmente reverter para um fenótipo virulento (GAO et al., 2013). Foi identificado que a proteína S2 pode interagir com duas isoformas da proteína de osteossarcoma (OS-9) equina, mas os papéis dessas proteínas celulares durante a infecção por VAIE não foram determinados (COVALEDA et al., 2010). A proteína S2 não é extensivamente conservada entre diferentes linhagens, com EIAV_{WY} compartilhando apenas 46,8%, 54,4% e 39,7% de identidade entre aminoácidos com EIAV_{IRE}, EIAV_{LIA} e EIAV_{MIY} respectivamente (COOK; LEROUX; ISSEL, 2013). No entanto, as substituições de aminoácidos parecem ser conservativas, e o motivo C terminal em S2 de todas as linhagens apresentam uma predominância de aminoácidos carregados positivamente (COOK; LEROUX; ISSEL, 2013).

1.2.5 Genoma de VAIE: long terminal repeats (LTR)

A região de repetição longa terminal (*long terminal repeat* - LTR) é responsável pela regulação da transcrição e poliadenilação do genoma viral, além de auxiliar na integração e transcrição reversa. O LTR pode ser dividido em três regiões designadas U3, R e U5. O LTR completo está presente apenas na forma proviral em ambas extremidades (Figura 3). Na forma livre de RNA viral, apenas a região R e U5 está presente na extremidade 5' do genoma e apenas a região U3 e R está presente na região 3' do genoma (MAURY, 1998). A região U5 de 50 nucleotídeos de comprimento em VAIE é mais curta do que a observada em muitos retrovírus (COOK; LEROUX; ISSEL, 2013). A região R é importante para a transcrição por conter o sinal para a cauda poli A e a TAR para ligar-se a *tat* (COOK; LEROUX; ISSEL, 2013). A região U3 contém a região promotora viral, que se liga a vários sítios de ligação do fator de transcrição, como o sítio de ligação PU.1, PU.2 e PEA1/AP1, bem como um sítio de proteína de ligação ao DNA metilado (*methylated DNA binding protein* - MDBP) e TATA box (CARVALHO; KIRKLAND; DERSE, 1993). A alta variabilidade da sequência de LTR em VAIE é observada principalmente em uma pequena porção do intensificador dentro desse segmento (MAURY, 1998). Estudos apontam que as pressões seletivas exercidas nas sequências potenciadoras de U3 em VAIE são diferentes das exercidas no restante da LTR,

mostrando que essa hipervariabilidade pode alterar a expressão do vírus em macrófagos teciduais e, portanto, contribuir para a doença clínica em equinos infectados (MAURY et al., 1997). Especula-se também que o tropismo celular de VAIE está relacionado a alterações no LTR e em *env*, que apresentam altos níveis de variação genética (MAURY et al., 2005). A alta variação do LTR foi observada em pôneis, quando após 112 dias de infecção experimental nenhum vírus com a sequência de LTR original foi encontrado (LICHTENSTEIN et al., 1999). O alinhamento das sequências de LTR demonstram que seu comprimento varia entre as linhagens de VAIE, com EIAV_{WY}, EIAV_{LIA}, EIAV_{IRE} e EIAV_{MIY} possuindo comprimentos de nucleotídeos de 323 pb, 316 pb, 313 pb e 306 pb, respectivamente. Além das diferenças de tamanho, há menos de 80% de identidade entre as sequências de LTR nas quatro linhagens diferentes (COOK; LEROUX; ISSEL, 2013).

1.3 Ciclo de replicação do vírus da Anemia Infecciosa Equina

Comparado ao vírus da imunodeficiência humana, um lentivírus muito estudado, muitos passos da replicação do vírus da Anemia Infecciosa Equina ainda precisam ser melhor elucidados. Uma característica unificadora dos lentivírus é sua replicação em macrófagos, mas em graus variados (FIELDS; KNIPE; HOWLEY, 2007). Os lentivírus primatas (HIV e SIV, por exemplo) replicam-se em macrófagos, mas a maioria dos vírus em um indivíduo infectado se origina de células CD4 + infectadas. Os lentivírus não-primatas são mais dependentes de macrófagos, exceto FIV e BIV, que se replicam bem em uma variedade de células (FIELDS; KNIPE; HOWLEY, 2007). O VAIE é capaz de infectar monócitos, mas o vírus se replica de forma muito mais eficiente quando a célula se diferencia em macrófago (MAURY, 1994). Dessa forma, o vírus usa os monócitos para se dispersar por todo o organismo e começa a se replicar em alto nível quando o monócito entra no tecido e se diferencia, uma característica observada em outros lentivírus (FIELDS; KNIPE; HOWLEY, 2007; GENDELMAN et al., 1986).

Quando se trata de replicação em retrovírus, podemos dividir as diferentes fases em estágios inicial e tardio. O estágio inicial contempla desde a entrada na célula hospedeira até a integração no genoma celular, e o estágio tardio compreende da expressão do RNA viral até a

montagem e brotamento das partículas virais imaturas (STAKE et al., 2013). No estágio inicial, a fusão do envelope viral com a célula hospedeira é promovida pela ligação da proteína de envelope com o receptor equino de lentivírus 1 (*equine lentivirus receptor 1* - ELR-1), presente em monócitos e macrófagos, embora um co-receptor desconhecido também possa estar envolvido (ZHANG et al., 2005). Essa entrada se dá por uma via endocítica dependente de pH baixo (BRINDLEY; MAURY, 2005; JIN et al., 2005) e mediada por clatrina (BRINDLEY et al., 2008). Após esse evento, a transcrição reversa começa utilizando a transcriptase reversa presente no vírion.

A transcrição reversa do genoma RNA viral para DNA de fita dupla é a marca característica dos retrovírus e deste passo deriva o nome do grupo. A reação ocorre em um grande complexo contendo proteínas codificadas por *gag*, RT, IN, e o RNA viral. A transcrição reversa normalmente começa logo após a entrada do núcleo do vírion no citoplasma da célula infectada, com o emparelhamento de um iniciador RNAt na região U3 do RNA viral denominada sítio de ligação de *primer* ou PBS (*primer binding site*) (Figura 4). O DNA é primeiro sintetizado a partir deste iniciador, utilizando o genoma de RNA de cadeia positiva como molde para formar sequências de DNA de cadeia negativa. A síntese ocorre em direção à extremidade 5' do RNA até CAP, gerando as sequências U5 e R. O primer RNAt permanece ligado à sua extremidade 5'. Esta síntese também pode ser iniciada “prematuramente” durante a montagem e libertação do vírion, de tal modo que as preparações de vírions podem conter pequenas quantidades dos intermediários iniciais do DNA proviral (FIELDS; KNIPE; HOWLEY, 2007).

A degradação do híbrido DNA:RNA é mediada pela atividade da RNase H da RT a fim de liberar a fita de DNA para ser transferida até a extremidade 3' do genoma (FIELDS; KNIPE; HOWLEY, 2007). A proteína NC auxilia na translocação do fragmento de DNA até a outra região R no genoma. Embora existam relatos que essa translocação ocorra para a região R do outro RNA genômico (translocação em trans), a melhor evidência é que a transferência ocorra aleatoriamente para qualquer uma das duas fitas de RNA viral (YU et al., 1998). Ao se ligar na extremidade 3', esse fragmento de DNA cria um iniciador para a síntese do DNA proviral pela RT, e a medida que o genoma entra na forma híbrida DNA:RNA pela ação da enzima, o RNA torna-se novamente suscetível à ação de RNase H da RT (FIELDS; KNIPE; HOWLEY, 2007). O *primer* responsável pela síntese da segunda fita de DNA é resultante da

digestão do RNA híbrido, pois uma região com sequência rica em purinas perto da extremidade 3' do genoma, o ppt, é relativamente resistente à atividade da RNase H. Com isso, a fita positiva é sintetizada até a extremidade 3' do genoma usando a fita negativa de DNA como molde. Em HIV já foi descrito mais de um ppt no genoma, criando sítios secundários de iniciação de cadeia positiva importantes para a conclusão adequada da transcrição reversa (CHARNEAU; ALIZON; CLAVEL, 1992). O *primer*, uma vez que serviu para iniciar a síntese de DNA, é rapidamente removido do DNA pela ação da RNase H (FIELDS; KNIPE; HOWLEY, 2007). A fita positiva é novamente transferida para a outra região R, de forma que a síntese da segunda fita se completa (Figura 4). O produto final da reação de transcrição reversa é uma versão de DNA linear de cadeia dupla de comprimento total do genoma, flanqueada em cada extremidade por cópias da LTR. Formas de DNA proviral circularizadas com a união do LTR podem ser encontradas também, embora não seja consenso que estas possam se integrar no DNA genômico. Este dsDNA forma o complexo pré-integração (*pre-integration complex* - PIC) juntamente com a integrase IN e outras proteínas virais e celulares, e este migra e entra no núcleo com o auxílio de S2 (LI et al., 2000).

A integração do DNA retroviral no genoma do hospedeiro é essencial para o ciclo de vida de retrovírus. Mutantes incapazes de integrar não estabelecem a infecção na célula (FIELDS; KNIPE; HOWLEY, 2007). Não há mecanismo para eliminar o provírus eficientemente uma vez que esteja estabelecido: este é passado para as células filhas durante a divisão celular e estabelece uma infecção persistente (FIELDS; KNIPE; HOWLEY, 2007). A integração é mediada pela IN, que ainda no PIC no citoplasma corta o dinucleotídeo TT das extremidades do DNA proviral, deixando as pontas do LTR com extremidades 3'-OH recuadas (Figura 7). Esta primeira reação catalisada pela IN é chamada de processamento. Dentro do núcleo, a integrase promove a clivagem do DNA alvo hospedeiro com um escalonamento de 4 a 6 pb, bem como a ligação das extremidades 3'-OH do DNA proviral a fosfatos alternados do alvo no DNA hospedeiro clivado. A remoção dos 5' nucleotídeos não complementares do DNA proviral e o reparo das lacunas no DNA do hospedeiro são realizados por enzimas do hospedeiro, que geram um provírus covalentemente integrado, mais curto em 2 pb em cada extremidade e flanqueado por duplicações do local alvo (ANDRAKE; SKALKA, 2015).

O genoma proviral integrado marca o fim da fase inicial de replicação do retrovírus, que é caracterizada pela ação das enzimas virais. Agora, o genoma proviral será processado por enzimas do hospedeiro, com etapas normais à célula, como a transcrição, *splicing* e tradução. Os transcritos retrovirais são modificados com a adição de uma cauda poli (A), como os RNAm celulares. Uma fração do RNA viral é processada e exportada para fora do núcleo para servir como RNAm para tradução em proteínas virais. O restante do RNA viral permanece sem ser processado e deve ser transportado do núcleo para o citoplasma, onde ele serve duas funções: (i) como o modelo para a tradução das proteínas estruturais virais, Gag e Gag-Pol; e (ii) como o genoma viral, que é empacotado em partículas virais. A proteína *rev* auxilia na exportação do RNAm do núcleo (STAKE et al., 2013). O RNA genômico forma um dímero não covalente (duas cópias) e é encapsulado por meio de uma interação entre a sequência de empacotamento e o NC (nucleocapsídeo). A proteína *tat* auxilia no aumento da taxa transcricional com a ligação a TAR, localizado no LTR (TANG; KUHEN; WONG-STAAAL, 1999). As glicoproteínas de *env* são inseridas na membrana plasmática via Golgi.

A montagem da partícula retroviral para o brotamento é impulsionada principalmente pela proteína precursora *gag*. O agrupamento de *gag* faz com que a membrana a envolva com as glicoproteínas virais e, eventualmente, cause o brotamento (STAKE et al., 2013). A localização do brotamento de lentivírus parece ser específica do vírus e da célula hospedeira. O sítio de montagem das partículas de VAIE ocorre na rede trans-Golgi, enquanto que em HIV-1 a montagem é direcionada para a membrana plasmática (ZHANG et al., 2016). Uma vez embalada, a protease viral cliva a poliproteína *gag/pol* para formar o viríon maduro (FIELDS; KNIPE; HOWLEY, 2007).

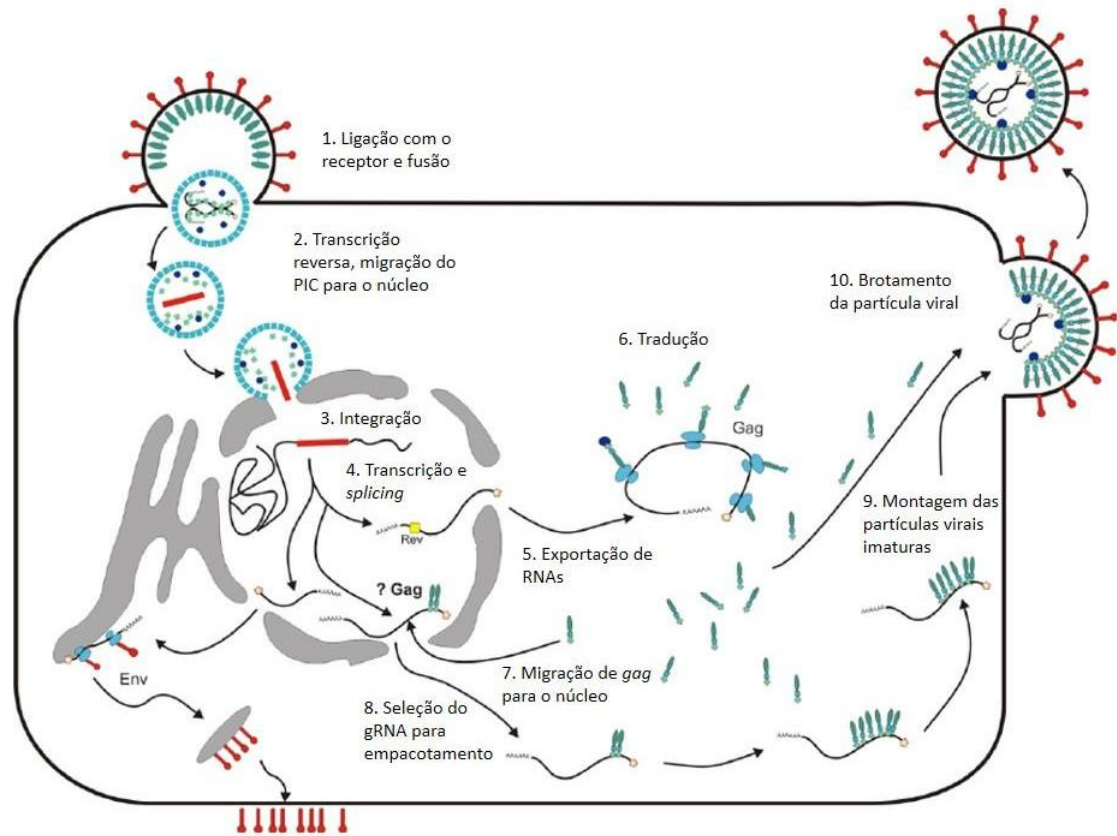


Figura 4: Modelo de replicação em retrovírus. A infecção inicia-se com a ligação da glicoproteína *env* viral ao receptor de superfície celular e a subsequente fusão do envelope viral com a membrana celular (1). O genoma viral é transcrito reversamente de RNA para DNA (2) e o provírus é integrado no genoma da célula hospedeira (3). A partir do genoma proviral, o RNA viral é transcrito pela polimerase celular e parte dele sofre *splicing* alternativo para traduzir a glicoproteína *env* e outras proteínas acessórias (4). Uma porção do RNA viral sem processamento é exportado do núcleo pela proteína *rev* (5) para traduzir *gag* e *gag-pol* (6). As proteínas *gag* de alguns retrovírus migram através do núcleo durante a montagem (7). Foi postulado que a população nuclear de *gag* (marcada com "?Gag") poderia selecionar o RNA genômico (RNAg) e transportá-lo para o citoplasma para o empacotamento. Alternativamente, a seleção do gRNA pode ocorrer no citoplasma (8). Em ambos os casos, o complexo *gag*-RNAg é transportado para a membrana plasmática (9) onde moléculas adicionais de *gag* se ligam para completar a montagem da partícula viral imatura, que brota a partir da membrana plasmática (10). Modificado de STAKE et al. (2013).

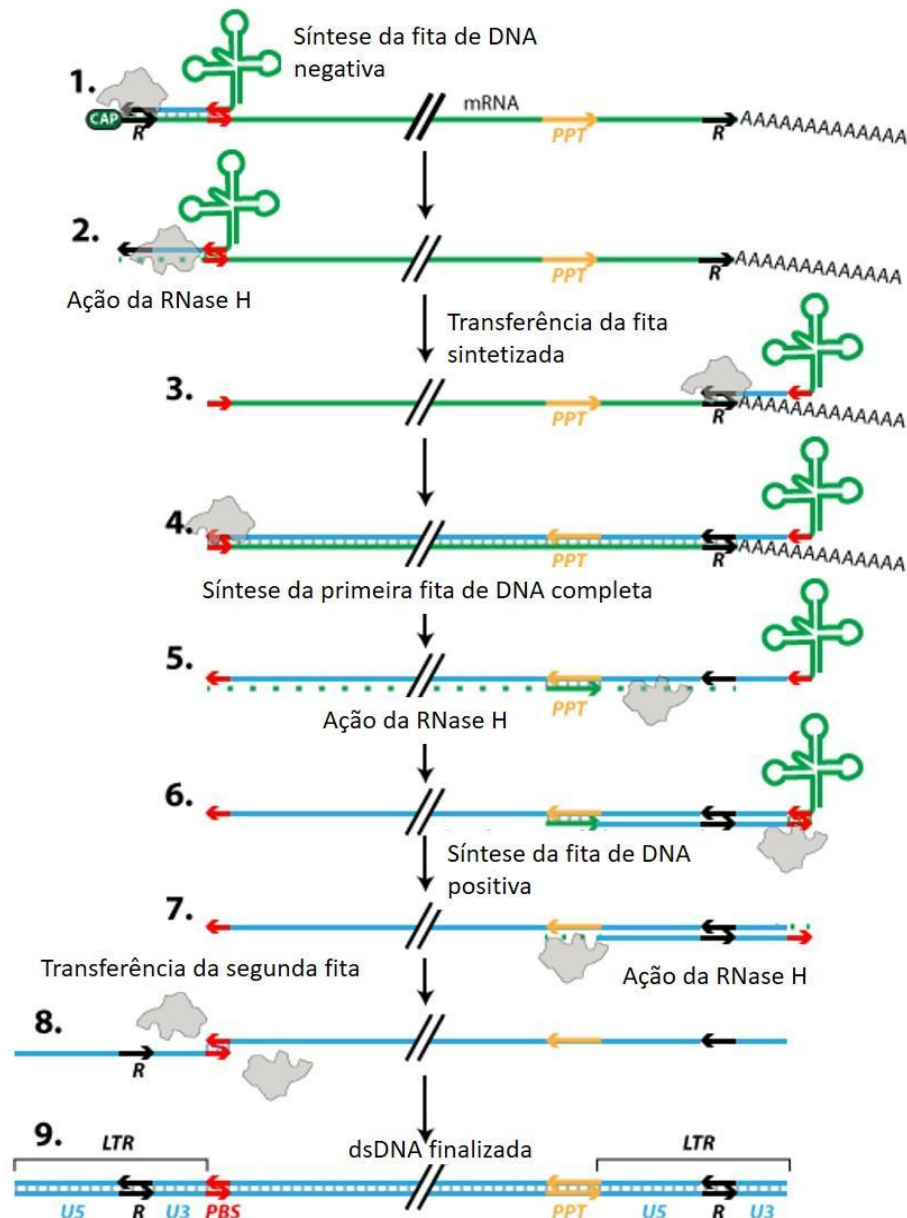


Figura 5: Esquema de incorporação do RNA genômico viral no genoma do hospedeiro como DNA proviral. (1) O RNAt celular liga-se ao sítio de ligação ao *primer* (PBS) do RNA genômico (RNAg) para iniciar a transcrição reversa até CAP. (2) RNase H viral degrada o RNA hibridizado com o DNA. (3) O DNA recém sintetizado é transferido e hibridizado na região de repetição do fim do genoma. (4) O genoma completo é copiado pela transcriptase reversa. (5) RNase H viral degrada o RNA hibridizado com o DNA, exceto a região ppt que resiste a degradação. (6) A síntese da segunda fita de DNA inicia-se no RNA hibridizado no ppt. (7) RNase H viral degrada o *primer* ppt e libera o RNAt ligado ao DNA. (8) O iniciador PBS liga-se a sua sequência complementar no DNA sintetizado. (9) As fitas positiva e negativa do DNA são sintetizadas para gerar o DNA proviral, que permanece associado às proteínas de capsídeo formando o complexo pré integração (pre-integration complex - PIC). Modificado de ViralZone (<https://viralzone.expasy.org/1937>). Acesso em 19/01/2019.

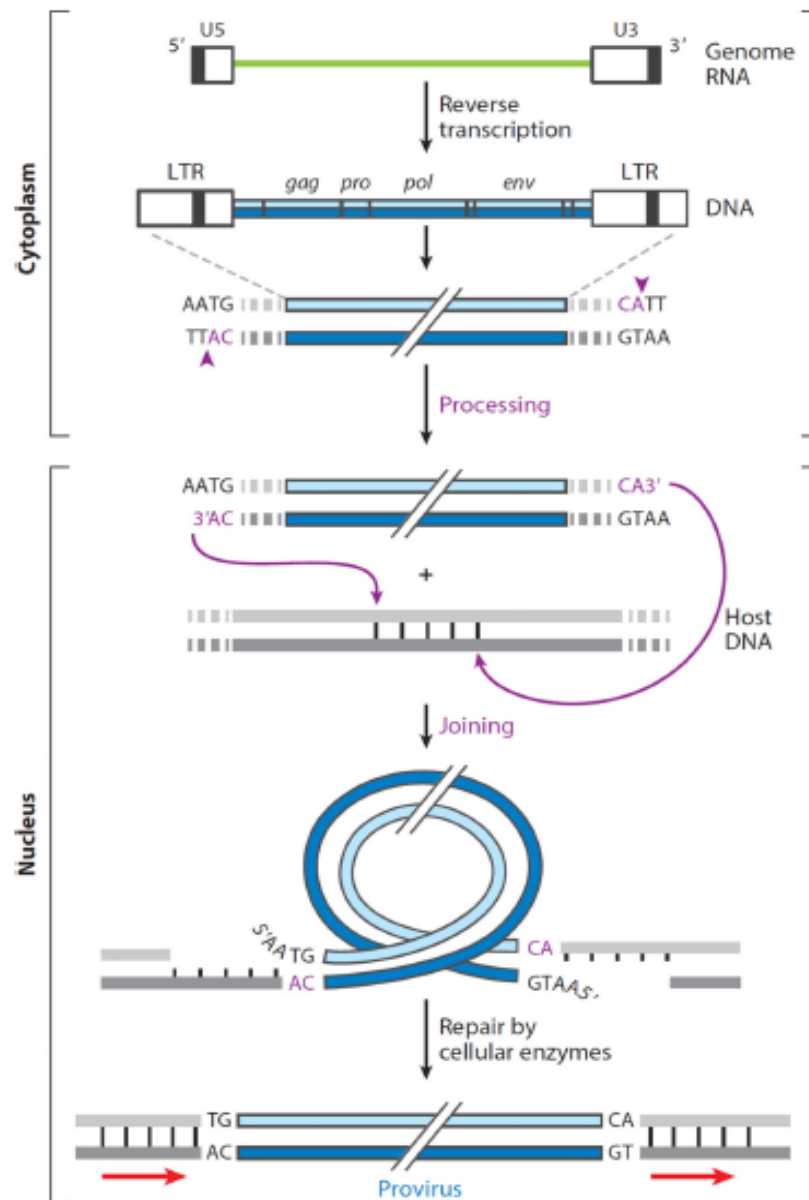


Figura 6: Integração do DNA proviral no genoma do hospedeiro. Modificado de ANDRAKE; SKALKA (2015).

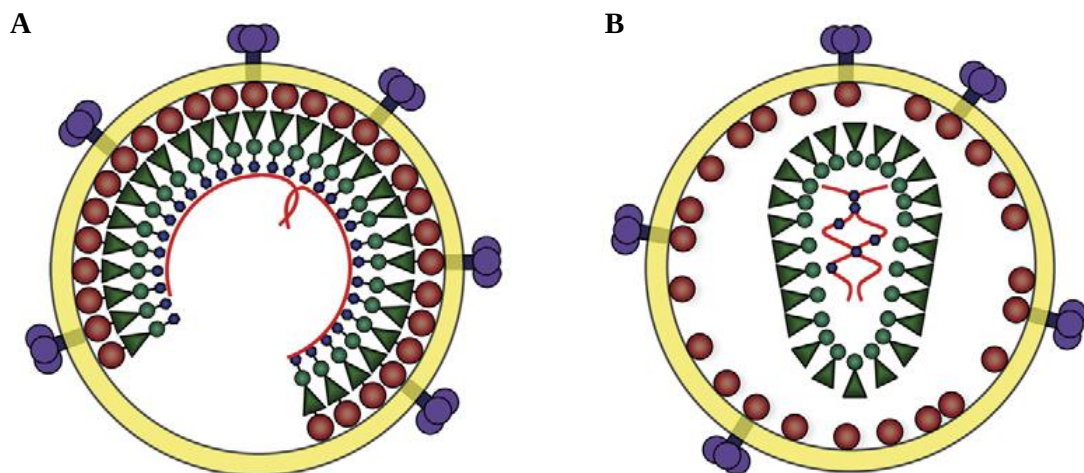


Figura 7: Modelo esquemático de vírion imaturo (A) e vírion maduro (B). As proteínas derivadas de cap são rearranjadas no vírion maduro: MA (círculo vermelho), CA (N terminal: triângulo verde; C terminal: círculo verde) e NC (círculo azul). Modificado de GANSER-PORNILLOS; YEAGER; SUNDQUIST, (2008).

1.4 Patogenia e clínica

Os sinais clínicos da doença são inicialmente causados por citocinas pró-inflamatórias que incluem o fator de necrose tumoral alfa ($\text{TNF}\alpha$), interleucina-1a e -b (IL-1a, IL-1b), interleucina 6 (IL-6) e fator beta transformação de crescimento ($\text{TGF}\beta$). Essas citocinas são liberadas quando cargas virais atingem um nível viral de 5×10^7 a 1×10^8 cópias/mL. A IL-6 e $\text{TNF}\alpha$ induzem respostas febris e aumentam a produção de prostaglandina E2, enquanto $\text{TNF}\alpha$ e $\text{TGF}\beta$ contribuem para a trombocitopenia e anemia através da supressão do crescimento de megacariócitos equinos e inibição da eritropoiese (REIS; LEITE, 2016). A anemia também é decorrente da hemólise de natureza imunológica, mediada por anticorpos específicos contra as glicoproteínas de superfície viral que se encontram ligadas aos eritrócitos, associada à fração C₃ do complemento (PERRYMAN et al., 1971). Além disso, em algumas espécies, $\text{TNF}\alpha$ induz trombocitopenia grave estimulando a liberação de agonistas de plaquetas incluindo trombina, plasmina e serotonina. A indução de estresse oxidativo produzido por alterações na glutathione peroxidase e nos níveis de ácido úrico

também podem desempenhar um papel na patogênese das infecções por VAIE através da cascata de respostas inflamatórias enquanto, simultaneamente, diminui a proliferação de células imunitárias (BOLFA et al., 2012). A hepatoesplenomegalia observada em necropsias é resultante do processo inflamatório desencadeado pela deposição de imunocomplexos e subsequente destruição de plaquetas e dos eritrócitos por macrófagos hepáticos (REIS; LEITE, 2016). Alguns animais podem apresentar leucopenia com discreta linfocitose e um aumento do número de monócitos circulantes (REIS; LEITE, 2016).

O período de incubação da doença e a clínica variam, dependendo da dose infectante, virulência e susceptibilidade individual do hospedeiro. Os primeiros sinais clínicos podem aparecer de três dias a quatro semanas após a infecção, sendo que já houve relatos de até 70 dias após a infecção (REIS; LEITE, 2016). A anemia infecciosa equina pode ter três fases clínicas distintas: aguda (episódio inicial da doença), crônica (episódios de sequenciais de viremia recorrentes) e inaparente (assintomático). No entanto, os sinais de doença podem estar completamente ausentes ou limitados a leves episódios febris que não são detectados. Os sinais clínicos mais comuns na fase aguda incluem febre, trombocitopenia e letargia. Em casos graves, petéquias, anemia hemolítica e epistaxe podem ocorrer. No início da fase aguda os animais podem ser soronegativos, fato que justifica a normativa do MAPA em pedir exames de animais em quarentena em 30 e 60 dias. Essa fase regride em alguns dias, podendo o animal manifestar uma forma grave e fatal da doença (menos de 5%) ou entrar na fase crônica ou na inaparente. Os sinais clássicos da doença crônica incluem anemia, trombocitopenia, perda de peso, edema e ocasionalmente sinais neurológicos (ataxia e/ou encefalite), e podem durar de meses até um ano. A maioria dos equídeos naturalmente infectados se encontra na fase assintomática ou inaparente, sem sinais clínicos e com baixa viremia. Esses animais quando não identificados mantêm o vírus nos plantéis, e em situações de estresse ou administração de fármacos imunossupressores, podem entrar em viremia depois de anos (ISSEL et al., 2014).

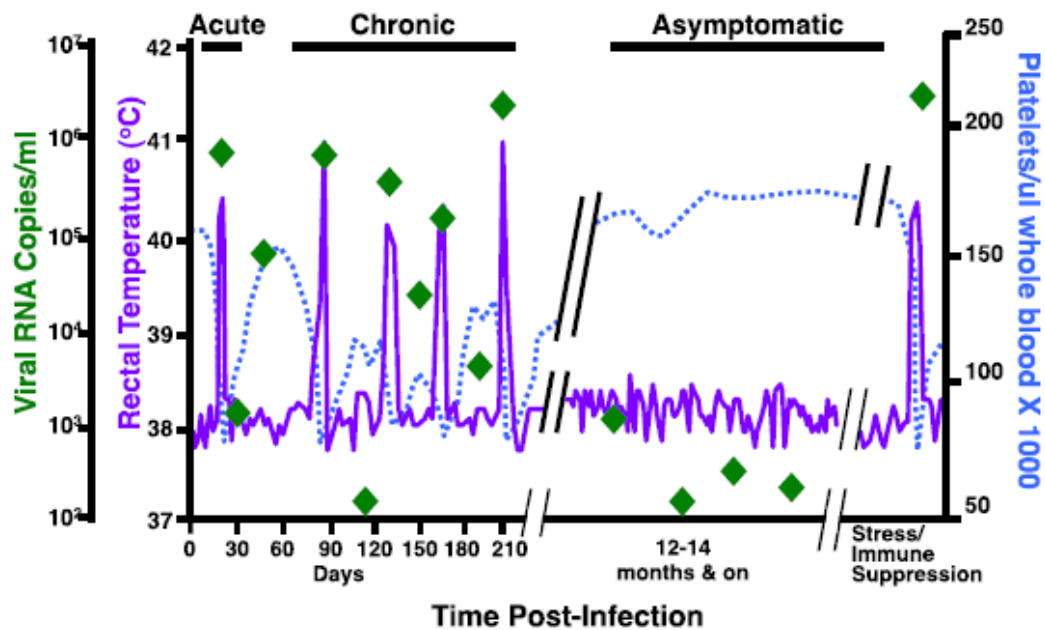


Figura 8: Perfil da infecção clínica pelo vírus da Anemia Infecciosa Equina. O diagrama esquemático apresenta quantificação viral no plasma, número de plaquetas no sangue e temperatura retal no decorrer dos estágios da doença. Episódios febris são caracterizados por temperatura acima de 39 °C, trombocitopenia é definida por níveis de plaquetas abaixo de 105.000 por uL de sangue e viremia é considerado acima de 10^5 cópias de RNA/mL de plasma. Fonte: CRAIGO; MONTELARO (2013).

Como observado em outros lentivírus, ainda não existe tratamento que consiga eliminar o vírus do hospedeiro. A terapia de suporte pode ser administrada para auxiliar na recuperação de episódios febris e sinais clínicos associados. O tratamento com corticosteróides é contra-indicado devido ao consequente aumento da carga viral (REIS; LEITE, 2016).

1.5 A transmissão da anemia infecciosa equina

A anemia infecciosa equina é uma doença de distribuição mundial cuja transmissão envolve a transferência de sangue do equídeo infectado para outro suscetível. A transmissão natural se dá por meio de insetos vetores (transmissão mecânica) ou da contaminação do filhote nascido de uma fêmea infectada (transmissão vertical) (REIS; LEITE, 2016). O vírus não se replica nos insetos vetores, apenas contaminando o aparato bucal durante uma alimentação interrompida. Os artrópodes relacionados com a transmissão são da ordem Diptera, como *Stomoxys calcitrans* e os gêneros *Tabanus*, *Chrysops* e *Hybromitra* (FOIL et al., 1983). O tamanho grande desses insetos permite carregar uma quantidade significativa do sangue infectado, o que o capacita como vetor mecânico. Dessa forma, esse meio de transmissão está limitado ao número e hábitos dos insetos hematófagos, a densidade dos animais, a interrupção da alimentação do inseto, forçando este a procurar outro animal em um curto período de tempo, e a quantidade de vírus circulante no equino infectado, do qual o inseto terá que se alimentar inicialmente (ISSEL et al., 2014).

Estudos revelam que o vírus permanece infectante de 30 minutos a 4 horas no inseto, e a distância máxima que os vetores realizam o repasto sanguíneo é 183 metros (REIS; LEITE, 2016). Experimentalmente foi determinado que 99% dos tabanídeos estudados retornarão ao hospedeiro original se uma alternativa está a mais de 50 metros (ISSEL; FOIL, 1984). Esses dados são importantes para determinar a distância de animais em quarentena, ou mesmo em fazendas com animais positivos, manter parte do plantel em regiões isoladas. Mosquitos são vetores altamente improváveis do vírus, pois além de não poder se replicar no vetor, o intestino do mosquito possui enzimas que impedem que ocorra a migração para as glândulas salivares. Como vetores mecânicos, a pequena quantidade de sangue que permanece no aparato bucal não é sequer comparável com a quantidade em um tabanídeo, o que seria insuficiente para uma transmissão (FOIL et al., 1987). Regiões tropicais e subtropicais do mundo todo são apontadas como locais de fácil difusão da doença, pois a população de insetos hematófagos vetores auxilia nessa transmissão. No entanto, em muitas regiões consideradas endêmicas, como o Pantanal brasileiro, apesar da presença dos vetores estes não são a principal via de transmissão da doença (ROSSI; TERNES, NOGUEIRA 2015).

Atualmente, a transmissão com intermédio humano por uso comum de materiais contaminados com sangue infectado, como equipamentos cirúrgicos, trocarte, grossa dentária, arreios, pinças de casco, tatuadores, esporas e outros fômites (utensílios que podem gerar

soluções de continuidade e extravasamento de sangue), principalmente agulhas, é o principal meio de transmissão em algumas localidades (REIS; LEITE, 2016). Nesses equipamentos, a capacidade infectante do vírus pode ser de até 96 horas e, no caso de agulhas, transfere um volume de cerca de 1.000 a 10.000 vezes maior de sangue infectado do que o inseto hematófago (ISSEL et al., 2014).

A transmissão também pode ocorrer verticalmente, pela placenta de éguas com alto título viral, com o contato pelo canal no parto, ou via colostro e leite para os potros. Os fetos de fêmeas positivas podem ser abortados ou nascerem já soropositivos. Entretanto, esse meio de transmissão não é importante epidemiologicamente em áreas controladas. O controle de obtenção de potros negativos a partir de fêmeas positivas é possível e já é empregado no Pantanal Brasileiro (SILVA et al. 2004).

1.6 Diagnóstico

O diagnóstico para anemia infecciosa equina não pode se basear apenas na manifestação clínica ou em achados patológicos, principalmente devido à presença dos animais com sintomatologia inaparente. As técnicas de diagnóstico atuais para AIE dependem da detecção de anticorpos específicos. Embora esta seja uma abordagem indireta e seja incapaz de detectar infecções recentes, a combinação dos dois principais testes recomendados pela Organização Mundial de Saúde Animal (World Organisation for Animal Health – OIE) permitem uma boa sensibilidade (ensaio de imunoadsorção enzimática - ELISA) associada a uma alta especificidade (teste de imunodifusão em gel de ágar - IDGA) (OIE, 2013). O isolamento viral não é utilizado em diagnóstico de AIE (OIE, 2013). Em muitos países, o único teste oficialmente reconhecido para o diagnóstico de infecções por VAIE é o IDGA (ISSEL et al., 2014).

O teste de imunodifusão em gel de ágar ou teste de Coggins detecta anticorpos contra a principal proteína do capsídeo de VAIE (p26) (COGGINS; NORCROSS; NUSBAUM, 1972). Os resultados deste teste provaram correlacionar-se positivamente com os resultados do teste de inoculação em cavalo usando 250 mL de sangue total, método diagnóstico

utilizado até a década de 1960 (PEARSON et al., 1971). O teste baseia-se na migração do antígeno e do anticorpo presente no soro do animal, formando uma linha de precipitação visível a olho nu. Inicialmente o antígeno para IDGA era extraído do baço de animais infectados, mas nos dias de hoje ele já é produzido por tecnologia recombinante (COOK; LEROUX; ISSEL, 2013). O IDGA é recomendado pela OIE e adotado como método principal de diagnóstico pelo MAPA (BRASIL, 2004; OIE, 2013), pois apresenta uma ótima especificidade. Em meados de 1980, testes sorológicos baseados em ELISA foram validados comparando resultados em amostras de referência e de campo com os do IDGA. O teste de ELISA apresenta maior sensibilidade que o IDGA, sendo que há muitos relatos de animais negativos no IDGA e positivos no ELISA (ISSEL et al., 2014; SCICLUNA et al., 2013).

O ensaio de imunoadsorção enzimática (ELISA do inglês *Enzyme Linked ImmunoSorbent Assay*) também é um ensaio sorológico que utiliza como antígeno a p26, ou uma combinação da p26 com a proteína do envelope gp45 (COOK; LEROUX; ISSEL, 2013; SUZUKI et al., 1982). Há quatro tipos de teste de ELISA disponíveis comercialmente e recomendados pela OIE: um competitivo e três não competitivos (OIE, 2013). Apenas o terceiro não competitivo usa p26 e gp45. Um positivo no ELISA deve ser retestado pelo IDGA e, em caso de resultados conflitantes, pode-se usar o imunoblot para averiguação (OIE, 2013). No Brasil, um quinto ELISA utilizando um gp90 recombinante (rpg90 ELISA) foi padronizado para corrigir possíveis falsos negativos e positivos resultantes do IDGA e ELISA comercial (REIS et al., 2012). O teste foi comparado em 1160 amostras de soro e demonstrou uma concordância de 95,2 % com IDGA. Atualmente, um novo kit comercial de ELISA foi lançado na Itália e utiliza como alvo uma proteína quimérica recombinante de gag e env (SCICLUNA et al., 2018). Este teste mostrou uma alta especificidade e um limite de detecção maior que o IDGA.

O teste de imunoblot tem sido usado desde o final da década de 1980 (ISSEL; COOK, 1993; ROSSMANITH; HORVATH, 1989) e pode detectar anticorpos contra a p26, e proteínas do envelope (gp90 e gp45). Inicialmente, o teste de imunoblot foi recomendado como teste complementar ao diagnóstico sorológico convencional porque amostras de VAIE em equídeos infectados com baixos níveis de anticorpos anti-p26 (que pode ser falsamente relatado como negativo no IDGA) geralmente apresentavam níveis mais altos de reatividade contra as proteínas do envelope. Esta descoberta levou a uma estratégia de combinação dos

pontos fortes dos testes vigentes: diagnóstico de ELISA primeiro, seguido pela confirmação pelo IDGA, com um teste de imunoblot caso os resultados de ELISA e IDGA não estivessem de acordo (COOK; LEROUX; ISSEL, 2013; ISSEL et al., 2013). O manual de diagnóstico da OIE sugere tal abordagem e, Itália e Estados Unidos a adotam (ISSEL et al., 2013; OIE, 2013; SCICLUNA et al., 2013).

Vários testes baseados na técnica de reação em cadeia de polimerase (PCR) foram desenvolvidos para detectar material genético de VAIE (diretamente do DNA proviral ou acoplado com um passo de transcrição reversa para o RNA viral), e alguns foram usados com sucesso com isolados de campo (CAPOMACCIO et al., 2012b; CAPPELLI et al., 2011; CURSINO et al., 2018; DONG et al., 2012; NAGARAJAN; SIMARD, 2001; QUINLIVAN; COOK; CULLINANE, 2007). As reações de PCR recomendadas pela OIE (somente como diagnóstico complementar ao sorológico) são uma Nested PCR do gene *gag* (NAGARAJAN; SIMARD, 2001) e uma qPCR precedida de transcrição reversa (RT-qPCR) (COOK et al., 2002). Tais oligonucleotídeos recomendados pela OIE falham em detectar algumas linhagens de VAIE de campo, como já descrito na Itália (CAPPELLI et al., 2011). Antes de estas técnicas poderem ser adotadas rotineiramente, deve ser demonstrado: (1) que os *primers* (e as sondas) utilizados nestes ensaios estão localizados em regiões altamente conservadas do genoma viral, uma vez que a variação nessas sequências pode impedir ou reduzir significativamente a sensibilidade diagnóstica; e (2) que as técnicas baseadas em PCR são sensíveis o suficiente para detectar níveis extremamente baixos de ácidos nucleicos específicos de VAIE presentes em alguns animais clinicamente inaparentes. Infelizmente, não há evidências conclusivas que demonstrem que qualquer um dos ensaios baseados em PCR descritos até a presente data atende a esses critérios. Assim sendo, as técnicas de diagnóstico atuais para AIE dependem ainda da detecção de anticorpos específicos para VAIE. Embora esta seja uma abordagem indireta e seja incapaz de detectar infecções recentes, a sorologia é a opção mais aceita no presente momento (ISSEL et al., 2014).

1.7 Evolução viral

A descoberta de um elemento retroviral endógeno “Foamy-like” no genoma de um celacanto (*Latimeria chalumnae*) sugere que os vírus da família *Retroviridae* têm evoluído com seus hospedeiros há pelo menos 400 milhões de anos (HAN; WOROBEY, 2012). O gênero *Lentivirus* é considerado recente em comparação com os demais retrovírus, sendo que a descoberta do lentivírus endógeno K (RELK) em coelhos e lebres demonstra que o gênero possui pelo menos 12 milhões de anos (KATZOURAKIS et al., 2007). No entanto, a gama de hospedeiros de lentivírus conhecidos é restrita a equídeos, bovinos, pequenos ruminantes, felinos e alguns primatas, e não ao longo da ordem filogenética de mamíferos. Recentemente, a descoberta de lentivírus endógenos em lagomorfos, colugos (primatas próximos) (HRON et al., 2014) e furão (*Mustelidae*) (CUI; HOLMES, 2012), sugere que os lentivírus podem ter uma distribuição mais ampla de hospedeiro (COOK; LEROUX; ISSEL, 2013).

Um mecanismo primário de resistência de lentivírus é a habilidade de evoluir em resposta a pressões do sistema imune do hospedeiro. É um paradigma na evolução desses vírus que a taxa de variação genética está correlacionada diretamente com os níveis de replicação viral e, conjuntamente, com a taxa de erro da polimerase do vírus. A transcriptase reversa (RT) de VAIE é tão propensa a erros como outras RT, pois pares errados de purina-pirimidina podem ser sintetizados pela enzima mais facilmente do que purina-purina. Este fato torna a fidelidade da enzima relacionada à sequência do genoma em questão (LEROUX; CADORÉ; MONTELARO, 2004). Como consequência da elevada taxa de erro, prevê-se que cada ciclo de transcrição reversa resulte na introdução de pelo menos uma mutação no genoma viral, produzindo aumentos significativos na diversidade genética. A organização genômica dos retrovírus, apresentando duas cópias de RNA viral, também permite uma recombinação genética, uma vez que variantes distintas podem ser empacotadas juntas, originadas de mutações pontuais ou coinfeção celular (SIMON-LORIERE; HOLMES, 2011). Os genes *gag* e *pol* são, no geral, os mais conservados do genoma, enquanto *env* e a ORF codificante para *rev* são as mais variáveis (COOK; LEROUX; ISSEL, 2013). As populações resultantes de vírus geneticamente relacionados competindo dentro de um ambiente altamente mutagênico foram denominadas "quasispécies". Um estudo envolvendo apenas um cavalo mostrou uma divergência de nucleotídeos de 2,6 % na gp90 em cinco episódios febris durante um ano com variação concentrada em sete regiões variáveis (LEROUX; ISSEL; MONTELARO, 1997). As LTRs, por outro lado, tendem a permanecer altamente conservadas em relação à cepa infectante, tanto no tempo como entre os

compartimentos dos tecidos (LEROUX; ISSEL; MONTELARO, 1997; MAURY et al., 2005; REIS et al., 2003). Por outro lado, o gene *rev* se desenvolve em duas subpopulações distintas, que exibem diferentes fenótipos em relação à virulência. Eles então circulam em dominância com cada estado de doença (BACCAM et al., 2003).

A alta conservação de p26 é aproveitada no fato de que este antígeno é a base para a maioria dos testes sorológicos disponíveis comercialmente para a detecção de equídeos infectados com VAIE (ISSEL et al., 2014). Estudos recentes sugerem que episódios cíclicos observados durante a fase crônica da AIE estão associados com populações virais distintas que podem ser distinguidas como variantes genômicas e/ou antigênicas. Estas observações sugerem que a natureza cíclica da fase crônica está ligada à produção e seleção de variantes de envelope viral (*env*), ou mutante de escape imunológico (LEROUX et al., 2001). Os mutantes de escape antigênico são vírus que por meio de mutações nas proteínas de superfície produzem infecções persistentes, permitindo a replicação em um hospedeiro mesmo na presença de anticorpos (FLORES, 2007). Essas conclusões são suportadas pela análise das sequências da proteína de superfície (SU – gp90) em cavalos e pôneis durante episódios clínicos cíclicos e nas substituições observadas em 8 domínios hipervariáveis (ZHENG et al., 1997). A análise da SU demonstra que as substituições de aminoácidos entre diferentes cepas estão distribuídas por toda a molécula com exceção do terminal amino. Outra característica conservada é a presença de resíduos de cisteína, sugerindo que as pontes dissulfeto são essenciais para a integridade estrutural e funcional da SU (ISSEL et al., 2014).

1.8 Genomas de VAIE sequenciados

Atualmente, sequências genômicas completas de VAIE isolados a campo em cinco países diferentes foram publicadas: EIAV_{LIA} (China) (WANG et al., 2002) (LIU et al., 2000), EIAV_{IRE} (Irlanda) (QUINLIVAN et al., 2013), EIAV_{MIY} (Japão) (DONG et al., 2013), EIAV_{WY} (Estados Unidos) (PETROPOULOS, 1997) e EIAV_{ITA} (Itália) (CAPPELLI et al., 2017). As quatro primeiras foram obtidas através da técnica de sequenciamento de Sanger (primeira geração). Apenas as últimas e mais recentes sequências – EIAV_{ITA} – foram obtidas com técnicas combinadas de sequenciamento de Sanger e sequenciamento massivo (ou *next*

generation sequencing – NGS). As sequências da China, Japão, Estados Unidos e Europa (Itália e Irlanda) compartilham aproximadamente 80% de identidade, formando quatro clados separados, sugerindo evoluções independentes de um ancestral comum (CAPOMACCIO et al., 2012a; CAPPELLI et al., 2011, 2017; ISSEL et al., 2014). Tal fato dificulta abordagens comuns a todas as regiões no que se trata de diagnóstico molecular e programas de desenvolvimento de vacinas (ISSEL et al., 2014).

Não há vacinas comerciais disponíveis contra VAIE atualmente. Uma vacina atenuada foi desenvolvida na China nos anos 70 pela passagem em série de uma cepa de vírus virulento através de culturas de leucócitos de burro (WANG et al., 2016). Esta vacina foi utilizada em cavalos chineses entre 1975 e 1990 e foi creditada como a responsável pelo controle de AIE na China, após o qual o programa de vacinas foi descontinuado (ISSEL et al., 2014). Embora as vacinas de VAIE atenuadas possam proteger contra viremia e doença clínica, as cepas de vacina podem persistir *in vivo* e não garantem que tais vírus não possam evoluir no hospedeiro. Devido à interferência com o teste de diagnóstico, além do potencial dos vírus para reverter a virulência, a vacinação não é um componente das estratégias de controle de AIE atuais (ISSEL et al., 2014).

1.9 VAIE no Pantanal, Brasil

O Pantanal é uma planície sazonalmente inundável, que ocupa 38,21% da área da bacia do Alto Paraguai com 138.183 km². Dezesesseis municípios compõem a área fisiográfica do Pantanal, sendo sete municípios no Estado de Mato Grosso, ocupando 35,36% da área do Pantanal, e nove municípios no Estado de Mato Grosso do Sul, ocupando 64,64% da área do Pantanal (SILVA; ABDON, 1998). O Pantanal é dividido em 11 sub-regiões e fazendas de criação extensiva de bovinos ocupam a maior parte da área. No último censo realizado foram contabilizados cerca de 9 milhões de bovinos e 124 mil cavalos (IBGE, 2016).

Historicamente, equídeos (cavalos, mulas e burros) têm sido essenciais à pecuária pantaneira. Baseada na utilização desses animais, a produção de gado de corte representa a atividade econômica de maior importância na região e encontra-se bem adaptada às condições

locais. A participação dos equinos tornou-se particularmente importante pela característica extensiva da atividade, assim como pelas difíceis e peculiares condições regionais, em função dos ciclos de cheia e seca (SILVA; ABREU; BARROS, 2001).

No Pantanal, animais soropositivos para VAIE são isolados e não eutanasiados, como nas outras regiões do Brasil (BRASIL, 2004). Segundo informações de fazendeiros e técnicos que vivem na região, a doença teria chegado em 1974 (SILVA; ABREU; BARROS, 2001). Hoje, os estados brasileiros do Mato Grosso e do Mato Grosso do Sul apresentam elevado número de casos reportados anualmente (Figura 9). Estes estados juntos representam de 10 a 40 % dos casos brasileiros, variando no ano analisado (Figura 10).

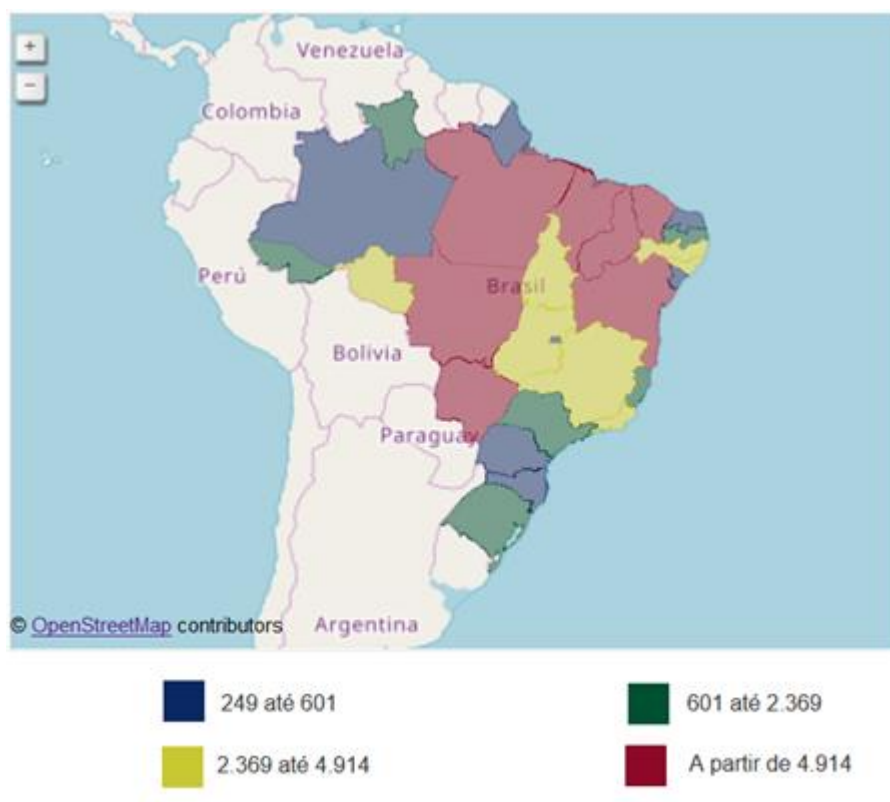


Figura 9: Casos de AIE notificados pelo Ministério da Agricultura, de 2005 a 2017 nos Estados brasileiros. Fonte: Base de dados do Sistema Nacional de Informação Zoossanitária - SIZ, Ministério da Agricultura, Pecuária e Abastecimento - MAPA (<http://indicadores.agricultura.gov.br/saudeanimal/index.htm>). Acesso em 01/08/2018.

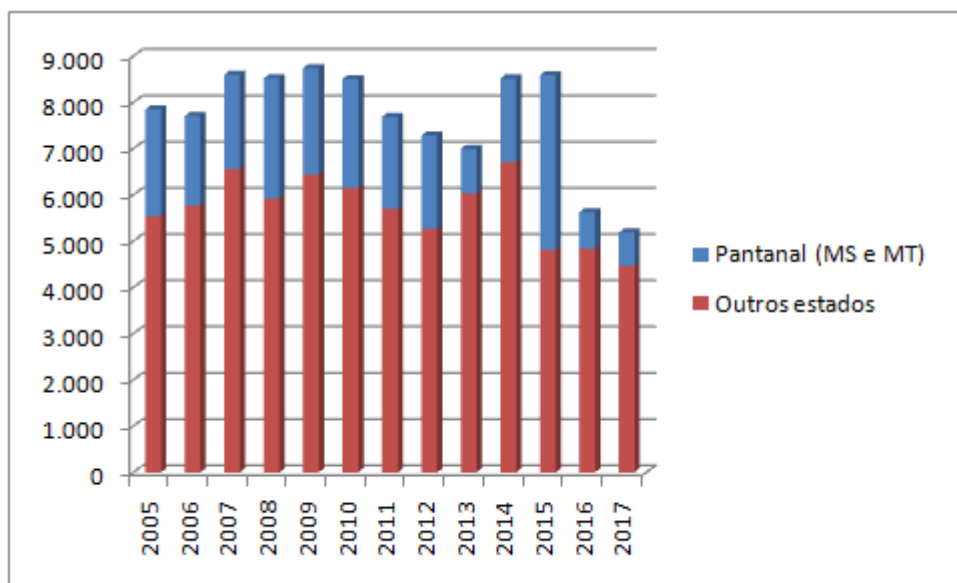


Figura 10: Comparativo entre o número de casos reportado no Pantanal Brasileiro e outros estados. Fonte: Base de dados do Sistema Nacional de Informação Zoossanitária - SIZ, Ministério da Agricultura, Pecuária e Abastecimento - MAPA (<http://indicadores.agricultura.gov.br/saudeanimal/index.htm>) Acesso em: 01/08/2018.

Um primeiro estudo realizado de 1990 a 1995 pela Embrapa Pantanal envolveu 3.285 cavalos pertencentes a 28 fazendas. Observou-se uma prevalência média de 24,8% da AIE nos equídeos estudados, e 18,17% em animais de trabalho. A prevalência foi significativamente maior nos animais que possuíam maior contato com o homem, isto é, mais manejados (SILVA et al., 1999a).

No mesmo ano, um estudo em uma fazenda na sub-região da Nhecolândia (MS) avaliou a prevalência da AIE em 252 cavalos de serviço e 16 chucros (sem doma). A soropositividade foi de 34,1% e 5,6% entre os animais de trabalho e os chucros, respectivamente, mostrando novamente prevalência significativamente mais elevada nos animais manejados pelo homem (SILVA et al., 1999b).

De setembro a novembro de 2009, outro estudo abrangeu 40 fazendas distribuídas na área de Corumbá (MS), e amostras de soro foram obtidas de 721 equinos e 232 muare. A prevalência da AIE nesta população aumentou de 18,17% para 38,60%, demonstrando a situação preocupante no Pantanal Brasileiro (NOGUEIRA et al., 2017).

Amostras de soro coletadas entre janeiro e julho de 2010 de 547 equídeos na região do Pantanal Norte (Poconé/MT) foram avaliadas com IDGA para o VAIE. A prevalência do vírus em fazendas foi de 52,0% (13/25) com prevalência ajustada nos equídeos de 31,5%. O estudo também demonstrou que fazendas e animais em áreas regularmente inundadas tinham respectivamente 60 e 146 vezes maior chance de serem soropositivos do que fazendas e animais localizados em áreas não inundadas. Apesar dessa associação, os dados apontam que outros fatores além das áreas inundadas, como a gestão de práticas, por exemplo, provavelmente desempenham um papel significativo na transmissão de VAIE (BORGES et al., 2013).

No Pantanal, já foram registradas 23 espécies de tabanídeos. A grande maioria dessas espécies ataca equídeos. Variações estacionais na umidade, pluviosidade e temperatura influenciam não apenas a fauna silvestre e a flora, mas também as populações de vetores e o surgimento de doenças nos animais. Em geral picos populacionais foram observados próximos ao início do período chuvoso (setembro-outubro), sendo o período até janeiro o de maior abundância de tabanídeos na região (BARROS; FOIL, 1999). Embora a maioria das espécies de mutucas ocorra durante todo o ano, sua maior abundância na primeira metade da época chuvosa sugere que este período seja o de maior risco de transmissão de agentes patogênicos (incluindo o VAIE) aos equídeos. Os riscos de transmissão nesse período também aumentam por causa do pico populacional de espécies de elevado potencial como vetor, tais como o *Tabanus importunus* (BARROS, 2001).

Em 2015, a estimativa da incidência de novos contágios foi calculada para o Pantanal Brasileiro através de modelagem matemática. Tal cálculo demonstra que o efeito iatrogênico tem o maior potencial de transmissão que os vetores insetos já descritos (ROSSI; TERNES, 2015).

A fim de reverter a situação encontrada na região, a Embrapa Pantanal juntamente com entidades governamentais criou o Programa de Prevenção e Controle da Anemia Infecciosa Equina no Pantanal do Estado do Mato Grosso do Sul (SILVA et al., 2004). O Programa preconiza o diagnóstico sorológico inicial e segregação dos animais positivos, bem como o uso de equipamentos de monta e doma separados. Também há o objetivo de obter potros negativos a partir de éguas positivas para uma gradativa redução (por mortalidade e

pela retirada dos potros) do número de animais positivos na fazenda, até a completa eliminação.

2 Justificativa

No Brasil, bem como em outros países, o diagnóstico da AIE é baseado em exames sorológicos para detecção de anticorpos virais. Entretanto, como exposto, os testes utilizados não são totalmente eficazes, havendo tanto a possibilidade de eutanásia de animais falso-positivos, quanto de manter animais positivos (falso-negativos) como fonte de infecção para outros animais. No Pantanal Brasileiro essa situação é mais crítica: a eutanásia de animais positivos não é obrigatória, sendo que a propriedade deve seguir rígidas normas de isolamento e manejo do plantel positivo totalmente separado do plantel negativo. O diagnóstico é principalmente baseado em IDGA, com o ELISA como teste secundário.

Tendo em vista a taxa evolutiva por mutações em lentivírus, uma região endêmica como o Pantanal Brasileiro, no qual os equinos vivem anos infectados com VAIE e posteriormente o transmitem para outros animais, deve possuir vírus circulante com características únicas. Considerando uma introdução do vírus em 1974 no Pantanal, há quase 50 anos o VAIE circula livremente em um número consideravelmente alto de equídeos, tempo suficiente para uma diferenciação dos demais vírus observados em outros países. A descoberta que sugere que o quadro clínico cíclico em equinos infectados provém de populações virais distintas emergentes durante a infecção nos dá maior segurança em dizer que animais positivos mantidos no plantel propiciam uma maior probabilidade de aparecimento de mutações e maior diversidade genética do vírus circulante na região.

A Semi Nested PCR para detecção de DNA proviral de VAIE proposta por Dong e colaboradores (2012) foi testada em estudos preliminares realizados no Laboratório de Diagnóstico Molecular, UNESP - Botucatu, para verificar sua eficácia e viabilidade para detecção de cepas locais provenientes da região do Pantanal. Obteve-se resultados satisfatórios em algumas amostras identificadas como positivas (IDGA e ELISA), mas muitas outras também positivas não amplificaram. Ao realizar o sequenciamento das amostras e

confirmar o material amplificado como DNA proviral do VAIE, observou-se uma variabilidade genética considerável das amostras entre si e quando comparadas com as sequências já depositadas em bancos de dados internacionais. Assim, podemos levantar a hipótese de que o vírus circulante no Pantanal possui sequência genética diferente o suficiente para não ser detectado por PCRs baseadas em sequências de outros países. Levando isso em consideração, bem como a alta variabilidade genética de retrovírus e o panorama único ao qual o VAIE é exposto no Pantanal, é possível que outras regiões virais utilizadas em diagnóstico (p26, gp45 e gp90) também apresentem mutações capazes de permitir falsos negativos nos testes sorológicos utilizados hoje (IDGA e ELISA).

Para auxílio destas questões, o sequenciamento massivo (NGS) pode ser uma ferramenta útil. Ele possui a capacidade de fornecer gigabases de sequências de uma maneira rápida e prática em relação aos métodos de sequenciamento tradicionais utilizando o método Sanger (MOLIKI; ROHWER; DUTILH, 2012). O NGS permite a caracterização detalhada de todo o genoma viral mesmo em presença de ácidos nucleicos do hospedeiro. Em particular, esta tecnologia tem sido usada para examinar populações virais dinâmicas que consistem em um grande número de quasispécies, em especial os retrovírus. A caracterização de populações virais na infecção precoce pelo HIV, a análise de padrões de evolução do HIV e a identificação de cepas de HIV super infectantes tem sido estudada por NGS (BUZON et al., 2011).

A caracterização molecular do vírus da AIE circulante no Pantanal, por meio de técnicas de sequenciamento do material genético viral completo, determinará sua filogenia e possibilitará o desenho de iniciadores específicos para reações em cadeia pela polimerase eficazes na identificação do agente em amostras biológicas. Tais conhecimentos elucidarão resultados discrepantes obtidos em diferentes técnicas sorológicas e constituirão base teórica fundamental para pesquisas futuras desta retrovirose.

3 Objetivos

Este trabalho tem dois objetivos principais:

1. Sequenciar e analisar o genoma completo do vírus da Anemia Infecciosa Equina em amostras de plasma de equinos naturalmente infectados da região do Pantanal Brasileiro;
2. Padronizar uma técnica de PCR quantitativa em tempo real (qPCR) para detecção do DNA proviral do vírus da Anemia Infecciosa Equina em amostras de sangue e soro de equinos naturalmente infectados no Pantanal, determinando sua sensibilidade e especificidade analítica e relativa, e analisar como obter melhor acurácia no diagnóstico da AIE.

4 Artigo Científico 1

High variations in *Equine infectious anemia virus* genome from naturally infected horses in Pantanal, Brazil: an endemic region case

Camila Dantas Malossi¹, Eduardo G. Fioratti², Jedson Ferreira Cardoso³, Angelo Jose Magro¹, Erna Kroon⁴, Daniel de Moura Aguiar⁵, Alice Mamede Costa Marque Borges⁵, Marcia Furlan Nogueira Tavares de Lima⁶, Leila Sabrina Ullmann¹, João Pessoa Araujo Junior^{1*}.

1. São Paulo State University (Unesp), Institute for Biotechnology, Botucatu, SP, Brazil.
2. Agrarian Sciences Institute (ICA), Vales do Jequitinhonha e Mucuri Federal University (UFVJM), Unaí, MG, Brazil.
3. Center for Technological Innovation, Evandro Chagas Institute, Ananindeua, PA, Brazil.
4. Virology Laboratory. Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil.
5. Virology and Rickettsiosis Laboratory, Mato Grosso Federal University, Cuiabá, MT, Brazil.
6. EMBRAPA Pantanal, Corumbá, MS, Brazil.

*Corresponding Author: João Pessoa Araujo Junior. Address: Alameda das Tecomarias, s/n, Chácara Capão Bonito, CEP: 18607-440, Botucatu, SP, Brazil, CEP: 18607-440. Phone No.: + 55 14 3880-0423 e-mail: joao.pessoa@unesp.br

Abstract

Equine infectious anemia virus (EIAV) is a persistent lentivirus that causes equine infectious anemia (EIA). In Brazil, EIAV is endemic in Pantanal region and euthanasia is not mandatory in these areas. All of the complete genomic sequences published from field viruses are from North America, Asia, and Europe, and proviral genomic sequences are available. Sequences of gag and LTR are currently available from the Brazilian virus. Thus, the present study aimed to sequence EIAV's genomic RNA for the first time in naturally infected horses from an endemic area of Brazil. RNA from naturally infected horses plasma was used for library preparation for next-generation sequencing (NGS), and gaps in the consensus viral sequence obtained were completed using Sanger sequencing methodology. Complete viral genome from two horses was obtained and annotated. Putative genes were analyzed and compared with the previously described showing conservation in gag and pol genes, and high variations in LTR and env. The p26 protein, used to target in diagnosis, presents amino acids alterations in some described epitopes. Brazilian genome sequences present a nucleotide identity of 88.6% between each other, and 75.8 to 77.3% with main field strains, like EIAV Liaoning, Wyoming, Ireland and Italy isolates. Furthermore, phylogenetic studies using EIAV sequences against known viral strains of EIAV strongly suggests this Brazilian strain comprise a separate monophyletic group. With these results, it is possible to better characterize the virus circulating in Brazil and to cope with the challenges of the EIAV diagnosis in Brazil, mainly in an endemic region where different measures of control have been applied.

Keywords: *Equine infectious anemia virus*, Pantanal Brazil, endemic region, RNA genome, equine, molecular characterization.

Background

Equine infectious anemia virus (EIAV) is a lentivirus, with an almost worldwide distribution, infecting equids, such as horses, mules, and donkeys. It causes a persistent infection characterized by recurring febrile episodes associated with viremia, fever, thrombocytopenia, and wasting clinical signs; most animals progress from a chronic stage characterized by recurring peaks of viremia and fever to an asymptomatic stage of infection. The inapparent carriers remain infective for life. Equine infectious anemia (EIA) is transmitted by biting insect vectors, mainly horseflies, and stable flies, as well as by contaminated syringes, needles, blood-contaminated instruments, and contaminated blood used in transfusions. No effective treatment or vaccine exists for the disease (COOK; LEROUX; ISSEL, 2013; ISSEL et al., 2014).

One of the main resistance mechanisms of lentivirus is its ability to evolve in the response of the host's immune system challenge. It is a paradigm in the evolution of these viruses that the mutation rate is directly correlated with the viral replication levels and, also, with the virus polymerase error rate. The EIAV reverse transcriptase (RT) is so prone to errors as any other RT, with wrong pairs of purin-pyrimidine occurring more often in extent than purine-purine. This fact makes the enzyme fidelity related to the current genome sequence (LEROUX; CADORÉ; MONTELARO, 2004). The *gag* and *pol* genes are, in general, are the most conserved genes of the genome, while *env* and *rev* are the most variable (COOK; LEROUX; ISSEL, 2013). The core protein p26, coding by *gag* gene, presents a high conservation rate, and it is useful considering that this antigen is the base for the majority of the serological tests commercially available for the detection of equids infected with EIAV (ISSEL et al., 2014).

Studies suggest that cycling episodes observed during EIA's chronic phase are associated with distinct viral populations that can be characterized as genomic and/or antigenic variants. These observations suggest that the nature of the chronic phase is linked to the production and selection of the viral envelope (*env*), or immunologic escape mutant (LEROUX et al., 2001). The antigenic escape mutants are viruses that cause persistent infections through mutations in the surface proteins, allowing the replication even in hosts

with antibodies (SCHWARTZ; NANDA; MEALEY, 2015). These conclusions are supported by the analysis of the surface protein (SU - gp90) in horses and ponies in cycle clinical episodes and in substitutions observed in 8 hypervariable domains (ZHENG et al., 1997). The SU analysis shows that substitutions in amino acid between different strains are distributed through all molecules except for the amino terminal. Another conserved characteristic is the presence of residues of cysteine, suggesting that the disulfide bonds are essential for the structural and functional integrity of the SU (ISSEL et al., 2014).

Genomic viral sequencing is used to study viral variants, quasispecies, and protein changes. Among the *Lentivirus*, we highlight the characterization of viral populations in early HIV infection, the analysis of HIV evolution patterns and the identification of super-infective HIV strains that have been studied by NGS (BUZÓN et al., 2011). Currently, complete genomic sequences of field-isolated EIAV in five different countries were published: EIAV_{WY} (United States of America) (PETROPOULOS, 1997), EIAV_{LIA} (China) (LIU et al., 2000), EIAV_{MIY} (Japan) (DONG et al., 2013), EIAV_{IRE} (Ireland) (QUINLIVAN et al., 2013), and EIAV_{ITA} (Italy) (CAPPELLI et al., 2017). Only the last and most recent sequences - EIAV_{ITA} – were obtained by Sanger Sequencing combined to massive sequencing (or next-generation sequencing - NGS), the other four complete genomic sequences were obtained by Sanger sequencing methodology. The sequences from China, Japan, United States and Europe (Italy and Ireland) share approximately 80% of identity, forming four separated clades, suggesting an independent evolution from a common ancestor (CAPOMACCIO et al., 2012a; ISSEL et al., 2014; CAPPELLI et al., 2017). This makes common approaches difficult to all regions when it comes to molecular diagnosis and vaccine development programs (ISSEL et al., 2014).

In Brazil, like other countries, EIA diagnosis is based on serological exams for viral antibodies detection: the diagnosis is primarily based on agar gel immunodiffusion (AGID), with enzyme-linked immunosorbent assay (ELISA) as a secondary test (OIE, 2013). In Brazilian Pantanal, this situation is critical because euthanasia of the positive animals is not mandatory, and the property must follow straight isolation rules and positive animal management completely separated from the negative animals. Studies in Pantanal showed a mean prevalence of 24.8% of EIA in equids and 18.17% in working animals from 1990 to 1995 (SILVA et al., 1999a), 34.1% and 5.6% among working animals and undomesticated

animals, specifically in the sub-region of Nhecolândia (MS) (SILVA et al., 1999b). The prevalence was significantly higher in the animals that had more contact with humans, that is, more frequently manipulated. In 2009, it was found that in the region of Corumbá (MS) - South Pantanal, the prevalence of the disease increased from 18.17% to 38.6% (NOGUEIRA et al., 2017), and the Poconé region (MT) North Pantanal - presented 31.5% of infected horses. The worrying situation in the Brazilian Pantanal is aggravated by the importance of the equines in the local livestock production, presenting restricted traffic for infected animals besides economic losses associated with the disease (SOUZA et al., 2018).

Considering the evolutionary rate due to mutations in lentivirus, an endemic region such as Brazilian Pantanal, in which the equid remains years infected with EIAV and keep transmitting to other animals, the circulating virus must present unique characteristics. Admitting a virus introduction in 1974 in Pantanal (SILVA; ABREU; BARROS, 2001), for almost 50 years the EIAV circulates freely in a considerable number of equines, enough time for differentiation of the other viruses observed in other countries (GIFFORD, 2012). The finding that suggests that the cyclic clinical picture in infected equines comes from distinct viral populations emerging during the infection gives us greater assurance that positive animals that remain in the herd (not euthanized) provide a higher probability of mutation and increased genetic diversity of the circulating virus in the region (CAPOMACCIO et al., 2012b).

In this study, we aimed the molecular characterization of the circulating EIAV in the Pantanal region, through sequencing techniques of the complete viral genetic material. Our results provide knowledge to elucidate discrepant results obtained in different serological techniques and will constitute a fundamental theoretical basis for future research on this retrovirus.

Material and Methods

1. Sample collection and RNA purification

Plasma of two horses (BRA1 and BRA2) from Poconé, MT, Brazil, positive on AGID (COGGINS; NORCROSS; NUSBAUM, 1972) and ELISA (REIS et al., 2012) was collected for this study. BRA1 was an asymptomatic viral carrier, and BRA2 presented edema of lower parts, high fever and hyperemic mucous membranes. Both animals' samples were tested in Semi Nested PCR (CURSINO et al., 2018) for infection confirmation. Plasma was previously centrifuged at 8,000 x g for 10 minutes, and 200 μ L of supernatant plus the pellet were used for RNA extraction with TRIZOL LS Reagent (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's instructions. The RNA samples were quantified by a commercial fluorometer (Qubit 2.0, Invitrogen, Carlsbad, CA, USA) with the RNA Assay kit (Invitrogen, Carlsbad, CA, USA), and assessments of 260/280 and 260/230 were measured with NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). RNA samples were then treated with DNase I (Sigma-Aldrich, Saint Louis, MO, USA), according to the manufacturer's instructions.

2. Library preparation and massive sequencing

The dsDNA library of BRA1 equine was synthesized from total RNA with Strand-Specific RNA Library Prep Kit (Agilent Technologies, Santa Clara, CA, USA), following the manufacturer's recommendations. Once ready, the library was quantified with the Illumina Library Quantification Kit (KAPA Biosystems, Wilmington, MA, USA). This kit consists of a quantitative PCR (qPCR) standardized to amplify the adapters present in the library's fragments. The qPCR product generated was subjected to horizontal electrophoresis in 1.5% agarose gel. The ideal library size should be between 200 and 500 bp.

The BRA2 sample was processed with a different commercial kit, based on a different methodology for library preparation. For this method cDNA was synthesized with Impron-II Reverse Transcriptase System (Promega, Madison, WI, USA) and random primers, following the manufacturer's recommendation, and the second DNA strand was synthesized by the following reaction: RNase H (5U/mL) (New England BioLabs, Ipswich, MA, USA), T4 DNA Polymerase (5 U/ μ L) (Thermo Fisher Scientific), T4 DNA ligase (5 U/ μ L), 10X T4 DNA Ligase Buffer (10 μ L) (Thermo Fisher Scientific), dNTP mix (10 mM) (Promega) and nuclease-free water to complete a final volume of 100 μ L reaction mix. The reaction was

incubated at 12 °C for 60 minutes, followed by another 60 minutes at 22°C (modified from OKAYAMA; BERG, 1982). The obtained double-strand DNA was purified with Agencourt AMPure XP magnetic beads (Agencourt Bioscience Corporation, Beverly, MA, USA) and quantified in Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) with the dsDNA High Sensitive kit. Then, the sample was diluted to 0,2 ng/μL for library preparation with Nextera XT DNA Library Preparation kit (Nextera XT, Illumina Inc, San Diego, CA, USA). The protocol was followed according to the manufacturer until the third step (PCR Cleaning). Libraries were quantified by KAPA Library Quantification Kit for the Illumina Sequencing platform (KAPA Biosystems), according to the KAPA Biosystems manual, and the other Nextera-XT steps were not performed (ULLMANN et al., 2015).

The library prepared with the StrandSpecific RNA Library Prep kit (BRA1) was sequenced with the NextSeq System (Illumina) using a NextSeq 550 System High-Output kit (1x75 cycles), and Nextera XT library (BRA2) was sequenced using a NextSeq Mid-Output kit (2x150 cycles) (Illumina).

3. Bioinformatics analysis

A *de novo* analysis was performed with IDBA-UD (PENG et al., 2012) and SPAdes programs (BANKEVICH et al., 2012). Geneious R8 v. 8.1.9 (KEARSE et al., 2012) was used to complete the bioinformatics analysis, mapping the contigs and high quality reads to a reference genome. EIAV genome sequences Liaoning (AF327877), Wyoming (AF033820), Miyazaki 2011-A (JX003263), IRE F2 (JX480631), and one incomplete sequence, ITA DE (KM247554) were aligned with NGS reads. Alignments and phylogenetic trees were built with MEGA 7. The sequences were aligned by MUSCLE algorithm and trees were constructed using the neighbor-joining method and the Tamura-Nei model was used to stipulate distances with Bootstrapping to 1,000 replicates.

4. Primers design and Sanger sequencing

Primers were designed for different regions of the EIAV genome, flanking products from 800pb to 1,300pb, based on a consensus sequence obtained by a massive sequencing to

cover possible gaps and also to validate the obtained sequence. Primers were designed using the Primer 3 Software (ROZEN; SKALETSKY, 2000) and analyzed by OligoAnalyzer 3.1, available online on the Integrated DNA Technologies - IDT site (<https://www.idtdna.com/calc/analyzer>) and by “primer Blast” software (www.ncbi.nlm.nih.gov) (Table 1).

Table 1: Primers to cover EIAV genome regions.

Primer	Sequence 5' - 3'	Genome position
AIE_GEN F 463-482	CCA GAG CAC AGG ARG ACA GG	463-482
AIE_GEN R 926-906	CCT CTG GGG GTT AAG GGT CGG	926-906
AIE_GEN R 1352-1333	TAA GGC TCT TTK GGS CCT TG	1352-1333
AIE_GEN F 1333-1352	CAA GGS CCM AAA GAG CCT TA	1333-1352
AIE_GEN R 1963-1943	TTA CTC CCA CAA ACT GCT CAG	1963-1943
AIE_GEN R 2273-2254	GCC ATT ACC AAT TGT GCC CC	2273-2254
AIE_GEN F 3878-3897	GCC AGG TCA CAA GGG CAT AT	3878-3897
AIE_GEN R 4540-4521	CCT GCA GGT CCA GAT CCY TG	4540-4521
AIE_GEN R 4974-4955	CCA TGG TGT TTG KCC YCC CA	4974-4955
AIE_GEN F 5112-5132	GGG TGA TGG TGC TGT AGT GGT	5112-5132
AIE_GEN F 5934-5959	GTG TAC AGA TAG TGA TCA TTG TCA AG	5934-5959
AIE_GEN R 6749-6727	GCC CGA GAA GTA ACA GGA AAA GG	6749-6727
AIE_GEN F 6823-6842	GCT ATT GCT GCT AGT GCY AC	6823-6842
AIE_GEN R 8000-7981	AGA TGT AGC TGG ATT TAR CG	8000-7981
AIE_GEN F 7751-7770	CAA AGC GAA GGA GGA AAC AT	7751-7770
AIE_GEN R 8277-8258	GGG ACT CAG ACC GCA GAA TC	8277-8258

The products of PCR were amplified in a reaction for 25uL with 2x GoTaq Green Master Mix (Promega), 0,3 mM of each primer, 5 uL of sample and nuclease-free water to complete the reaction. The conditions of the first PCR cycles consisted of a start denaturation for 10 minutes at 95° C, 5 repeated cycles of 30 seconds at 95°C, 30 seconds at 60°C, and 2 minutes at 72°C, 5 repeated cycles of 30 seconds at 95°C, 30 seconds at 55°C, and 2 minutes

at 72°C, 30 repeated cycles of 30 seconds at 95°C, 30 seconds at 50°C, and 2 minutes at 72°C, followed by a 10 minutes final extension at 72°C. The PCR amplification was analyzed by horizontal electrophoresis in 1.5% agarose gel and purified for Sanger sequencing. Internal primers were designed to enhance the Sanger sequencing in large products. The amplify product of *env* gene region (nt 5112 to 8000; see Table 1) could not be sequenced only with PCR purification, demanding to clone in *E.coli* DH5- α with a pGEM-T Easy vector (Promega), following the manufacturer's protocol. For the PCR product sequencing, we used the kit BigDye Terminator v3.1 Cycle Sequencing (Life Technologies) with a reaction per primer, following the manufacturer's orientation. The capillary electrophoresis and reading were made in an automatic sequencing system ABI 3500 (Life Technologies).

5. Variant analysis

A search for variations and SNPs was performed using all the aligned reads to the final consensus BRA1 sequence obtained. Only regions with at least 1,000 reads coverage and 0.25 variance of frequency were analyzed using Geneious R8 v. 8.1.9 program tools.

6. Molecular modeling: p26 analysis

A molecular model of the gag protein p26 was built to verify possible variations of epitopes since such protein is used for the serological diagnosis (AGID). This molecular model was based on the 2.7 Å resolution-crystallographic structure of the p26 (1EIA) from a previously X-ray diffraction experiment (JIN et al., 1999). Linear p26 amino acid sequences from BRA1 and BRA2 were analyzed in HHpred program to search homology for threading (SÖDING; BIEGERT; LUPAS, 2005). Modeller 9.15 was used to produce models of protein tertiary structures (ŠALI et al., 1995). The best model was determined by geometry validation in MOLProbity (CHEN et al., 2010). Epitope sequences previously described (CHONG et al., 1991; SOUTULLO et al., 2007; HU et al., 2016) were compared with a reference model and obtained models with a structure alignment using CHIMERA (PETTERSEN et al., 2004).

Results

1. Complete sequence of Brazilian EIAV

The massive sequencing using the Illumina platform was performed with two samples of RNA (or cDNA) obtained from plasma samples of two naturally infected different horses. The sequencing of the equine 1 (BRA1) was obtained using the Strand-Specific RNA Library Prep Kit (Agilent Technologies) and yielded more reads than the sequencing of the equine 2 (BRA2), prepared with Nextera XT kit (Illumina) (Table 2). These were aligned using four complete genomes obtained from field animals - Liaoning (AF327877), Wyoming (AF033820), Miyazaki2011-A (JX003263), IRE F2 (JX480631) - and one incomplete genome also from a field animal, ITA DE (KM247554), as a reference.

Table 2: Summary statistics for the two equine infectious anemia virus sequenced genomes.

Samples	Raw reads	Reads at least Q30	Mapped reads	Mean coverage
BRA1	18,371,626	92.7%	422,676	3,969
BRA2	10,374,696	71.2%	6,927	78.3

All consensus of the alignments with the database references had some gap in different regions of the viral genome, especially in *env* gene.

In order to cover the remaining regions and to obtain the complete genomic sequence, primers were designed from the partial consensus BRA1 and BRA2 sequences and their sequenced products. The PCR product corresponding to the *env* gene did not show multiple bands at the gel electrophoresis, however, the sequencing indicated that more than one product had been amplified. In this case, cloning of the products was performed and many individual colonies were sequenced.

The sequences obtained by Sanger were incorporated to the map to reference analysis with each complete genome already described. Such sequences work as a guide in highly divergent regions of the original sequence and help on the genome coverage. Lastly, the consensus sequence obtained to BRA1 has 8,016 bases from R to R (LTR) and BRA2 has 7,968 bases, ending with a poly A tail, characteristic of the retroviruses. The consensus sequences were annotated with prediction ORFs, based in others genome annotation.

The multiple alignments of the complete sequences using MUSCLE were performed to generate a phylogenetic tree by neighbor-joining. Identities between the described genomes and the Brazilian obtained can be observed in Table 3. The sequences with higher identity among themselves are from Europe, with identities varying between 98.22 and 99.42%. Genomes from different regions show identity around 74 to 79%. Both Brazilian sequences, BRA1 and BRA2, have identity of 88.56%. The phylogenetic tree showed clades well delimited of the viral sequences of each region (Europe, China, Japan, United States of America and Brazil) (Figure 1). Together with a field sequence described in China, two sequences from cellular culture propagation sequences used for the viral vaccine from China aligned in the same clade (HM141910 e HM141922).

Table 3: Identity matrix with EIAV genomes shows the percentage of bases/residues with are identical.

	Miyazaki	BRA1	BRA2	Wyoming	IRE H3	IRE F4	IRE F2	IRE F3	ITA SA	ITA DE	DLV9-7	DV10-3	Liaoning
Miyazaki- JX003263													
BRA1	74,384												
BRA2	74,281	88,596											
Wyoming - AF033820	75,432	76,257	75,953										
IRE H3 - JX480634	74,871	76,416	75,901	77,794									
IRE F4 - JX480633	74,919	76,476	75,946	77,854	99,226								
IRE F2 - JX480631	74,725	76,372	76,101	77,915	98,864	99,009							
IRE F3 - JX480632	74,988	76,340	75,814	77,794	98,985	99,190	99,420						
ITA SA - KM247555	74,729	76,403	76,259	77,891	98,273	98,515	99,033	98,819					
ITA DE - KM247554	74,528	76,264	76,196	77,828	98,223	98,388	98,969	98,781	99,324				
DLV9-7 - HM141922	76,139	76,127	76,354	78,162	77,891	77,923	77,887	77,803	78,058	77,995			
DV10-3 - HM141910	77,226	77,036	77,272	78,750	78,974	79,006	79,019	78,957	78,872	78,720	96,441		
Liaoning - AF327877	77,029	77,108	77,368	78,871	79,058	79,090	79,079	78,994	78,897	78,784	96,526	97,791	

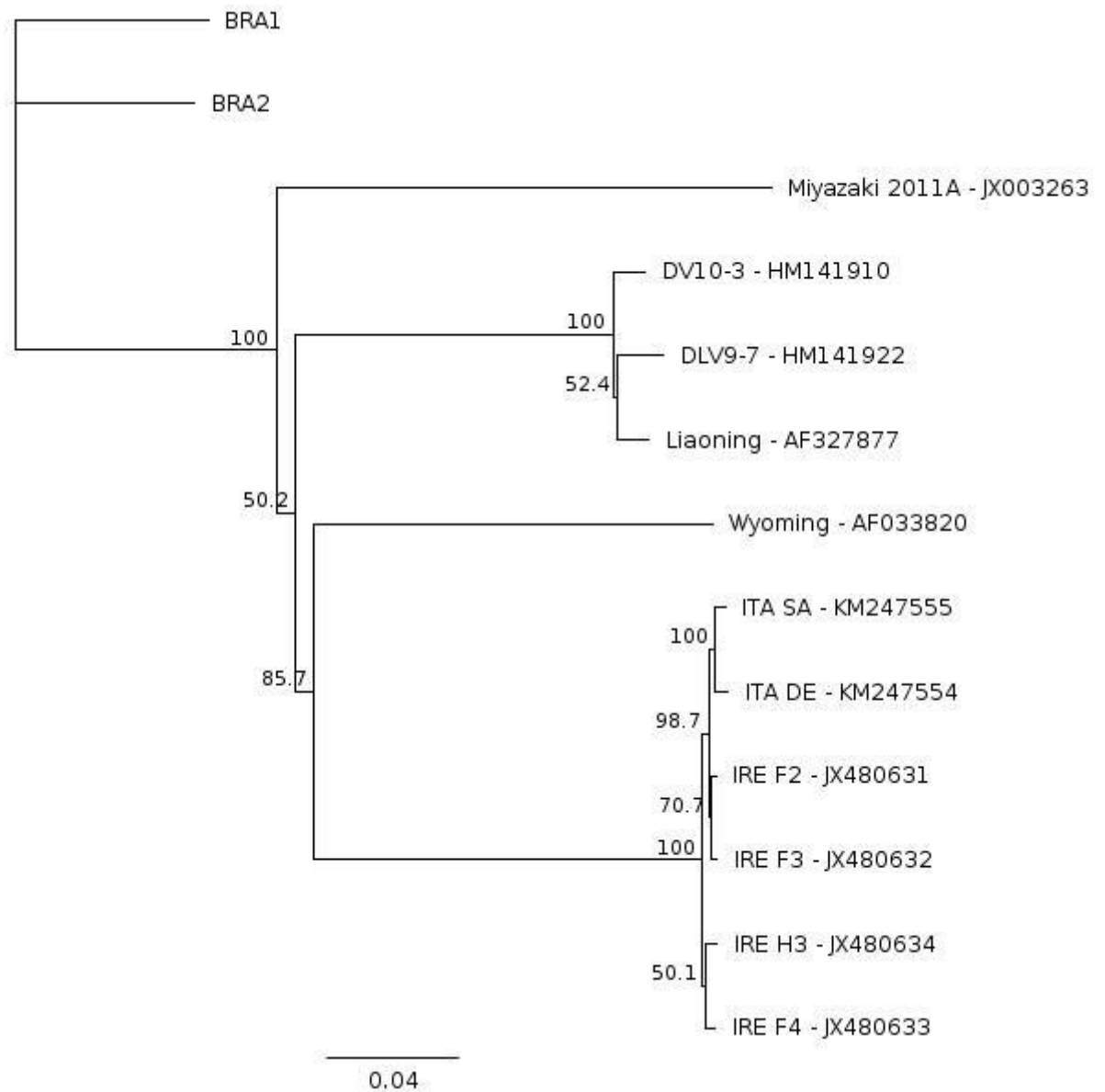


Figure 1: Cladogram involving EIAV complete genome sequences and the sequences obtained from BRA1 and BRA2 from Pantanal. Used sequences: EIAV Miyazaki2011A (Japan); EIAV BRA1 and EIAV BRA2 (Brazil); EIAV Liaoning, EIAV DLV9-7 and EIAV DV10-3 (China); EIAV Wyoming (USA); EIAV ITA SA, EIAV ITA DE, EIAV IRE F2, EIAV IRE F3, EIAV IRE F4 and EIAV IRE H3 (Europe).

2. Polymorphisms on EIAV genome

Genetic variation between the viral genomes was in the form of single nucleotide polymorphisms (SNPs), insertions and deletions. Only BRA1 had enough coverage for variation analysis (coverage above 1,000). In total, 33 SNPs were identified between the BRA1 reads, with 19 generating an amino acid change (Appendix A). The *pol* gene presents the highest number of SNPs (21), followed by *env* (7) and *gag* (4). Only one SNP is located in U5 LTR (Figure 2). No insertions and deletions passed in cut-off of 0.25 frequency.

Sanger cloning sequences of *env* were assembled to compare variations in this gene. Sequences from the same equine show insertions and deletions from three to 24 amino acids (1 to 6 amino acids), besides the nucleotides / amino acids substitutions (Appendix B). There is not one single sequence covering the total gene, but PCRs fragments to complete the whole gene sequence since Sanger sequencing presents a maximum length of 800-900 bp and the gene is higher than this. Unfortunately, we can not say if a clone assemble corresponds to the real sequence or if it is a chimera, but it demonstrates the variability in this gene.

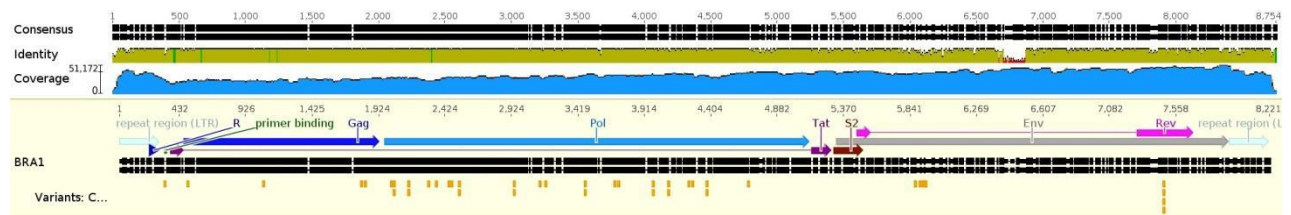


Figure 2: Representative genome of EIAV BRA1. Genome annotation shows the variants distribution, marked below in yellow. Coverage is represented in blue, and identity in dark yellow.

3. LTR and viral genes

The genomes of Brazilian EIAV do not present a complete LTR like proviral genomes. BRA1 and BRA2 sequences contain 192 bases in U3, 76 bases in R, 39 bases in U5 and 191 bases in U3, 78 bases in R, and 40 bases in U5, respectively. In total, it will be a complete BRA1 LTR (U3+R+U5) with 307 bases, and a 309 bases BRA2 LTR. Brazilian sequences of LTR were aligned with EIAV LTR from EIAV field full genome sequences and Chinese sequences (WANG et al., 2018). BRA1 and BRA2 LTR demonstrate 92.6 % identity

among each other, with variation in identity between 61.91 and 72.92 % with the foreign sequences. The most of the variation concentrates in U3 region, and the initiation site of transcription is changed from GGG/A to AGA in Brazilian sequences. Additional conserved motifs were found: a methylated DNA-binding protein site (MDBP, also called EF-C or EP) with variations (BRA1 = GTTGCTAGGCAAC, BRA2 = GATGATAAGCAAC), an absent PEA1/AP-1 site in both LTR, only two PU.1 binding sites (one site in Wyoming strain and three in another), TATA box and the transactivation response element (TAR) without alterations (Appendix C). An LTR phylogenetic analysis was performed to compare with full genome one. Each monophyletic grouping comprises the same sequences' region, with BRA1 and BRA2 together in a clade (Appendix G).

Both BRA1 and BRA2 *gag* genes present 1,464 nucleotides and encode 487 amino acids of Gag polyprotein (124 aa of p15, 232 aa of p26, 79 aa of p11, and 52 aa of p9). When compared with the world sequences, Brazilian *gag* sequences (BRA1 and BRA2) show identity between 77 and 79 %, and 91.12 % among each other. However, there are more amino acid variations between the Brazilian sequences (7 amino acids) than between the European ones (1 amino acid), that has approximate identity of 99 % among themselves. The leucine-rich type nuclear export signal is conserved in all p15 proteins and major homology region (MHR) of p26 presents the same sequence than Liaoning strain (IRQGPKEPYPEFVDRLLSQI). The first p11 zinc-binding domain is conserved for all amino acids and the second one presents a Q -> E alteration, as observed in Miyazaki strain. The YPDL late domain in p9 presents no variations (Appendix D).

The *pol* is the most conserved polyprotein in retrovirus, and all sequences analyzed show identity between 79.8 and 99.8 %. The Brazilian *pol* sequences present 3,414 nucleotides, and 1,137 amino acids in polyprotein, that is cleaved into five proteins: protease (PR), reverse transcriptase (RT), RNase H (RN), dUTPase (DU), and integrase (IN). BRA1 and BRA2 show 91.747 % identity. The protease amino acids that form the substrate binding site, DTG catalytic motif and active site flap were all conserved through all field strains and Brazilian ones (Appendix E). Only Miyazaki and Wyoming presents an alteration of valine (GIVG) for leucine or isoleucine, all nonpolar amino acids. The RT motifs described in HIV were also conserved in BRA1 and BRA2 (DIGD and DD) (DONG et al., 2014). The RN contains four invariant residues of one nucleotide, and a glycine-rich motif is conserved in all

sequences. Out of the five conserved motifs described in DU, Brazilian sequences show alterations in amino acids on M5 (BRA1 = RGEEGFGSTG; BRA2 = RGEKGFSGTG); the conserved residues in M3, that is responsible for binding the uracil ring (N67, G69, Y75, Q80, and I82), and the residue D72 is conserved in all strains. In the case of IN, the zinc-binding HHCC motif (H12, H16, C40, C43), hydrophobic core residues (I5, A8, L22, A33, and I36), and three central catalytic core residues (D64, D116, and E152) present no alteration.

EIAV *env* sequence encoded the surface protein (gp90) and transmembrane protein (gp45). BRA1 *env* presents 2,613 nucleotides and 872 amino acids; BRA2 *env* presents 2,568 nt and 855 aa. This is the most variable EIAV gene, with 72.83 % identity between BRA1 and BRA2 sequences, and at least 57.33 % of identity with another field strains. The gp90 protein variations were identified in 8 hypervariable regions (V1 to V8) and many mutation changes were observed (Appendix F). In fact, only the first 12 amino acids from gp90 form a conserved region in all the analyzed sequences. The cysteine residues are conserved. There are more variations between Brazilian sequences than European ones. The gp45 protein presents a transmembrane domain, with alteration of one amino acid between BRA1 and BRA2.

Phylogenetic trees were built based on the alignment of each analyzed gene: *gag*, *env*, and *pol* (Appendix G). All of them present clades well delimited of the viral sequences from each region (Europe, China, Japan, United States of America and Brazil). Together with the field sequence from China, both sequences propagated in cell culture used for vaccine studies (DLV9-7 and DV10-3) were aligned. Due to the high variability, the tree for *env* was built through the alignment of the nucleotides and also the amino acids, and both presented similar grouping. The phylogenetic analysis of the EIAV genes corroborated with the results observed on the LTR and full genome tree.

4. Gag variation and p26 molecular modeling

Brazilian p26 presents 232 amino acids, where p26 BRA1 presents the alteration in positions 73 (N -> L), 77 (I -> M), 80 (E -> D), 199 (K -> R), 210 (T -> S) and 215 (M -> L), and p26 BRA2 presents the alteration in position 73 (N -> L), 80 (E -> D) and 215 (M -> L). Comparing with previously described epitopes for p26, BRA1 and BRA2 have amino acids

alterations in ⁷³NLDKIAEE⁸¹ and ¹⁹⁹KNAMRHLRPEDTLEEKMYAC²¹⁸ (HU et al., 2016), and ²¹⁵MYACRD²²⁰ (SOUTULLO et al., 2007) (Table 4). The modified amino acid characteristics were not altered (hydrophobic, hydrophilic, etc). Antigenic recognition might be performed by other unmodified epitopes if any of these are not binding. The primary sequences of p26 BRA1 and BRA2 were modeled based on 1EIA (JIN et al., 1999) for structural comparison and alignment for epitopes' visualization. The structure of the epitopes and the modified amino acids on the surface are observed on them (Appendix H).

Table 4: Previously described epitopes alterations for each p26 sequence obtained. Alterations are marked in red.

Reference	Epitope	BRA1	BRA2
Hu et al. (2016)	⁷³ NLDKIAEE ⁸⁰	⁷³ LLDKMAED ⁸⁰	⁷³ LLDKIAED ⁸⁰
Hu et al. (2016)	¹⁹⁹ KNAMRHLRPEDTLEEK MYAC ²¹⁸	¹⁹⁹ RNAMRHLRPEDSLE EKLYAC ²¹⁸	¹⁹⁹ KNAMRHLRPEDTLE EKLYAC ²¹⁸
Chong et al. (1991)	¹⁵⁸ KEPYPEFVDRLLSQI ¹⁷²	¹⁵⁸ KEPYPEFVDRLLSQI ₁₇₂	¹⁵⁸ KEPYPEFVDRLLSQI ¹⁷²
Soutullo et al. (2007)	²⁰⁰ NAMRHL ²⁰⁵	²⁰⁰ NAMRHL ²⁰⁵	²⁰⁰ NAMRHL ²⁰⁵
Soutullo et al. (2007)	²¹⁵ MYACRD ²²⁰	²¹⁵ LYACRD ²²⁰	²¹⁵ LYACRD ²²⁰

Discussion

This is the first report of a complete genome of EIAV currently circulating in Brazil. Partial sequences of Brazilian EIAV *gag* gene from Bahia, Brazil, are already available on NCBI data bank, however without a reference paper (GenBank Accession: KC213783 - KC213790). This Brazilian State is not considered an endemic area for EIA, thus euthanasia of positive animals is mandatory, as for the other regions of the country with the exception of Pantanal. A partial sequence of 5' LTR plus a tat fragment with a Semi Nested PCR was amplified from equids from Pantanal, and they present variable nucleotide similarities ranging from 93.5% to 100% (CURSINO et al., 2018). Capomaccio and collaborators (2012) mention two Brazilian sequences but no data or specific origin is available.

Field full sequences from other countries (except Wyoming) were described from proviral DNA samples, and not from genomic RNA. Complete field sequences described in USA, Japan, China, Ireland, and a partial Italian sequence were compared with the new Brazilian sequences. While EIAV_{LIA}, EIAV_{IRE}, EIAV_{MIY}, and EIAV_{WY} strains were sequenced by Sanger method, Next Generation Sequencing (NGS) was used to obtain a partial sequence of EIAV_{ITA}, from the amplification of 7,841 bp of EIAV (CAPPELLI et al., 2017). In the present study, tRNA from equines sera were prepared to NGS with library preparation kits, without previously amplification.

Comparing StrandSpecific RNA Library Prep kit and Nextera XT kit, the first shows 1,8 times more reads than the second, this difference is more pronounced when we verify the percentage of these reads that correspond to the virus: from 1.9% to 2.1% in Agilent kit against 0.04% to 0.18% in Illumina Kit. This event may be related to the library preparation methodology employed in each protocol. The most efficient kit is based on a messenger RNA purification through poly A tail (also present in lentivirus genome), with a subsequent synthesis of cDNA, followed by library preparation. The Nextera XT kit works with tagmentation (transposases) and subsequent indexing of the total double-stranded DNA sample, and it should already have the prior preparation of cDNA and second strand synthesis before initiating the protocol. Alternatives for sample enrichment and better viral recovery for the Nextera XT protocol were also tested. We did not succeed in improving the RNA purification with different kits and protocols tested, prior enzymatic treatment for host DNA/RNA degradation, serial centrifugation cycles to isolate fractions with only viruses present, cDNA synthesis with oligo dT primer, and/or viral capture with specific primers (data not shown).

It was not possible to complete the viral sequence only with the reads obtained in the NGS. In order to close the gaps, PCRs using primers designed from BRA1 and BRA2 incomplete sequences amplified regions of the viral genome for Sanger sequencing. The PCRs designed to amplify fragments of the *env* gene could not be sequenced by Sanger only with the PCR product purification.

After cloning, different viral sequences were observed in the same equine, with a high-quality rate in the electropherogram. This level of variation is common in lentiviruses, in which the viral RT promotes insertion of errors because it does not possess a proofreading

mechanism. This leads to a gradual accumulation of different populations of viruses that differ slightly in their RNA sequence, resulting in a benefit in the face of natural selection (LEROUX et al., 2001). The *env* gene is the most variable in lentiviruses, with mutations in the gp90 being one of the responsible for the generation of immune system escape mutants (QUINLIVAN et al., 2013). Due to this high variability, the alignment algorithms presented difficulty in performing a map to reference that would provide a reliable consensus for the region. In fact, the *env* gene presented low identity among the aligned reads (Figure 2), the presence of SNPs (NGS data), insertions and deletions (Sanger data) and hypervariable regions V1 to V8 with many alterations when compared to the reference sequence (HOWE et al., 2002).

Studies of EIAV genomic variation during persistent infection in experimentally infected equid verified that febrile peaks and reemergence of symptoms are clearly associated with envelope sequences that modify antigenic properties of the virus, evidently as a result of immune screening. The predominant site of EIAV variation during persistent infection is gp90 (CRAIGO; MONTELARO, 2010). Variations of gp90 are identified in 8 hypervariable regions (V1 to V8). Through these regions, it is possible to differentiate quasispecies, with each presenting distinct signatures, especially in the V3 and V4 domains (HOWE et al., 2002). In BRA1 V3 region was found 6 SNPs (5881G>A; 5885C>G; 5912A>T; 5929C>A; 5941A>C; 5943C>T), and some nucleotides alterations, insertions and deletions in BRA1 clones (Appendix B). The V4 is a region shorter than V3, and it did not present predicted SNPs for the region, but insertion was found among the sequenced clones of BRA2 (Appendix B). The presence of two distinct signatures in the *env* sequencing shows that both equines were infected by different viral quasispecies. This fact is expected for an endemic region, in which animals live in infected years. In positive farms, not only natural selection generates new quasispecies through the escape of the host immune system, but there is also the same environment living of infected animals, allowing the exchange of circulating viruses in different equines, either by vectors action, either through human (AUTORINO et al., 2016).

Still, in *env* gene, the glycoprotein gp45 (transmembrane protein) is a smaller protein with less variation than gp90. For this reason, it is used in some ELISA tests as complementary antigen to p26 (ISSEL et al., 2014; SCICLUNA et al., 2018). In fact, the

Brazilian sequences of gp45 presented regions with more than 10 amino acids identical to those observed in other strains, only one SNP with substitution of the amino acid in gp45 (region shared with the *rev* gene) and insertion of 3nt (1aa).

Pol gene is described as a conserved sequence. In fact, there is a high identity between Brazilian sequences and references. It is the gene with more SNPs found. Furthermore, the fact a majority of these were non-synonymous, suggests strong individual host-mediated selective pressure. Quasispecies analysis of other EIAV strains shows this profile too (CAPPELLI et al., 2017). Protein domains are conserved, ensuring functional proteins to complete viral replication cycle.

The gag polyprotein produces four proteins (Matrix p15, Capsid p26, Nucleocapsid p11, and the Late protein p9) by proteolytic cleavage. Of those, p26 is important for generating the immune response detected by the gold standard diagnostic method, the AGID (OIE, 2013). Such a gene is chosen for diagnosis because it's well conserved in the different studied lineages and also due to its high antigenicity. Besides AGID, the PCRs suggested by OIE are also based on this gene amplification. Alterations on the nucleotides sequence that result in protein structural changes may influence in serological diagnosis if protein structure is changed. In fact, serologically silent equines have been observed in several regions of the world (ISSEL et al., 2013; SCICLUNA et al., 2013; RICOTTI et al., 2016). Both Brazilian p26 proteins show surface amino acids alterations in described epitopes (CHONG et al., 1991; SOUTULLO et al., 2007; HU et al., 2016). Studies to viral genes characterization to use as serological antigens or direct diagnosis target help in diagnosis development with better detection and specificity. These data contribute for disease control in an endemic local as Brazilian Pantanal, where viral evolution happens enable infected animals living for years with EIAV.

Comparative nucleotide sequence analysis provides important information about the extent of variation between geographically distinct EIAV strains. Here, we show two complete sequences of EIAV genome in Brazil. These samples are from an endemic region, and we find different viral quasispecies in each equine host analyzed. Comparison of LTR and structural gene sequences between the novel EIAV_{BRA1} and EIAV_{BRA2}, and field viruses show different monophyletic groups. The viruses that are currently sequenced are separated in local groups, indicating that EIAV could be separated into groups similar to HIV. The HIV

classification is divided into groups and subtypes according to origin and phylogeny, and it allows to trace outbreaks (HEMELAAR, 2012). There are not enough EIAV sequences to classify as HIV, so this study represents a significant step forward in increasing understanding of molecular conservation and variation in this important equine lentivirus.

Conclusions

Complete genomic sequences of Brazilian EIAV were successfully described from an endemic area. The NGS strategy developed was able to get enough reads for a full sequence, without pre-amplification, of the RNA equine sample. Based on sequence alignments and similarity plot analysis, BRA1 and BRA2 are not closely related to any other strain reported to date, forming a separate monophyletic group and as such can be classified as a novel EIAV strain. With *env* gene analysis, different viral quasispecies could be observed in both equines. Studies for variations on EIAV genome can help to understand the viral complexity and evolution mechanisms of EIAV in persistent and natural infection present in an endemic area.

References

- AUTORINO, G. L. et al. Evolution of equine infectious anaemia in naturally infected mules with different serological reactivity patterns prior and after immune suppression. **Veterinary Microbiology**, v. 189, p. 15–23, 2016.
- BANKEVICH, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. **Journal of computational biology : a journal of computational molecular cell biology**, v. 19, n. 5, p. 455–77, 2012.
- BUZÓN, M. J. et al. Deep molecular characterization of HIV-1 dynamics under suppressive haart. **PLoS Pathogens**, v. 7, n. 10, p. 1–11, 2011.
- CAPOMACCIO, S. et al. Geographic structuring of global EIAV isolates: A single origin for New World strains? **Virus Research**, v. 163, n. 2, p. 656–659, 2012a.
- CAPOMACCIO, S. et al. Detection, molecular characterization and phylogenetic analysis of full-length equine infectious anemia (EIAV) gag genes isolated from Shackleford Banks wild horses. **Veterinary Microbiology**, v. 157, n. 3–4, p. 320–332, 2012b.
- CAPPELLI, K. et al. Deep sequencing and variant analysis of an Italian pathogenic field strain of equine infectious anaemia virus. **Transboundary and Emerging Diseases**, v. 64, n. 6, p. 2104–2112, 2017.
- CHEN, V. B. et al. MolProbity: All-atom structure validation for macromolecular crystallography. **Acta Crystallographica Section D: Biological Crystallography**, v. 66, n. 1, p. 12–21, 2010.
- CHONG, Y. H. et al. Characterization of the antigenic domains of the major core protein (p26) of equine infectious anemia virus. **J Virol**, v. 65, n. 0022–538X (Print), p. 1007–1012, 1991.
- COGGINS, L.; NORCROSS, N. L.; NUSBAUM, S. R. Diagnosis of equine infectious anemia by immunodiffusion test. **American Journal of Veterinary Research**, v. 33, n. 1, p. 11–18, 1972.
- COOK, R. F.; LEROUX, C.; ISSEL, C. J. Equine infectious anemia and equine infectious anemia virus in 2013: A review. **Veterinary Microbiology**, v. 167, n. 1–2, p. 181–204, 2013.
- CRAIGO, J. K.; MONTELARO, R. C. EIAV envelope diversity: shaping viral persistence and encumbering vaccine efficacy. **Curr HIV Res**, v. 8, n. 1873–4251 (Electronic), p. 81–86, 2010.
- CURSINO, A. E. et al. Equine infectious anemia virus in naturally infected horses from the Brazilian Pantanal. **Archives of Virology**, v. 163, n. 9, p. 2385–2394, 2018.
- DONG, J. B. et al. Identification of a novel equine infectious anemia virus field strain isolated from feral horses in southern Japan. **Journal of General Virology**, v. 94, n. PART2, p. 360–

365, 2013.

GIFFORD, R. J. Viral evolution in deep time: Lentiviruses and mammals. **Trends in Genetics**, v. 28, n. 2, p. 89–100, 2012.

HEMELAAR, J. The origin and diversity of the HIV-1 pandemic. **Trends in Molecular Medicine**, v. 18, n. 3, p. 182–192, 2012.

HOWE, L. et al. Equine infectious anemia virus envelope evolution in vivo during persistent infection progressively increases resistance to in vitro serum antibody neutralization as a dominant phenotype. **Journal of Virology**, v. 76, n. 21, p. 10588–97, 2002.

HU, Z. et al. Identification and characterization of a common B-cell epitope on EIAV capsid proteins. **Applied Microbiology and Biotechnology**, v. 100, n. 24, p. 10531–10542, 2016.

ISSEL, C. J. et al. Challenges and proposed solutions for more accurate serological diagnosis of equine infectious anaemia. **Veterinary Record**, v. 172, n. 8, p. 210, 2013.

ISSEL, C. J. et al. Equine infectious anemia in 2014: Live with it or eradicate it? **Veterinary Clinics of North America - Equine Practice**, v. 30, n. 3, p. 561–577, 2014.

JIN, Z. et al. Model for lentivirus capsid core assembly based on crystal dimers of EIAV p26. **Journal of Molecular Biology**, v. 286, n. 1, p. 83–93, 1999.

KEARSE, M. et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, v. 28, n. 12, p. 1647–1649, 2012.

LEROUX, C. et al. Equine Infectious Anemia Virus Genomic Evolution in Progressor and Nonprogressor Ponies. **Journal of Virology**, v. 75, n. 10, p. 4570–4583, 2001.

LEROUX, C.; CADORÉ, J.-L.; MONTELARO, R. C. Equine Infectious Anemia Virus (EIAV): what has HIV's country cousin got to tell us? **Vet Res**, v. 35, p. 485–512, 2004.

LIU, H. et al. Complete Sequence of Proviral DNA of Equine Infectious Anemia Virus Strain L(Liaoning). **Chinese Journal of Preventive Veterinary Medicine**. S1, 2000.

NOGUEIRA, M. F. et al. Equine infectious anaemia in equids of Southern Pantanal, Brazil: Seroprevalence and evaluation of the adoption of a control programme. **Pesquisa Veterinaria Brasileira**, v. 37, n. 3, p. 227–233, 2017.

OIE. World Organization for Animal Health. Equine Infectious Anaemia. In: Manual of diagnostic tests and vaccines for terrestrial animals. Paris: OIE, 2013. p.1-6. Available in: <http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.05.06_EIA.pdf>. Access in 28 out. 2017.

OKAYAMA, H.; BERG, P. High-efficiency cloning of full-length cDNA. **Molecular and cellular biology**, v. 2, n. 2, p. 161–170, 1982.

PENG, Y. et al. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. **Bioinformatics**, v. 28, n. 11, p. 1420–1428, 2012.

PETROPOULOS, C. J. Retroviral taxonomy, protein structures, sequences, and genetic maps. In: COFFIN, J. M. (Ed.). **Retroviruses**. New York: Cold Spring Harbor Laboratory Press, 1997.

PETTERSEN, E. F. et al. UCSF Chimera - A visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605–1612, 2004.

QUINLIVAN, M. et al. Genetic characterization by composite sequence analysis of a new pathogenic field strain of equine infectious anemia virus from the 2006 outbreak in Ireland. **Journal of General Virology**, v. 94, n. PART3, p. 612–622, 2013.

REIS, J. K. P. et al. Recombinant envelope protein (rgp90) ELISA for equine infectious anemia virus provides comparable results to the agar gel immunodiffusion. **Journal of Virological Methods**, v. 180, n. 1–2, p. 62–67, 2012.

RICOTTI, S. et al. Serologically silent, occult equine infectious anemia virus (EIAV) infections in horses. **Veterinary Microbiology**, v. 187, p. 41–49, 2016.

ROZEN, S.; SKALETSKY, H. Primer3 on the WWW for General Users and for Biologist Programmers. **Methods Mol Biol**, v. 132, p. 365–386, 2000.

ŠALI, A. et al. Evaluation of comparative protein modeling by MODELLER. **Proteins: Structure, Function, and Bioinformatics**, v. 23, n. 3, p. 318–326, 1995.

SCHWARTZ, E. J.; NANDA, S.; MEALEY, R. H. Antibody Escape Kinetics of Equine Infectious Anemia Virus Infection of Horses. **Journal of Virology**, v. 89, n. 13, p. 6945–6951, 2015.

SCICLUNA, M. T. et al. Is a diagnostic system based exclusively on agar gel immunodiffusion adequate for controlling the spread of equine infectious anaemia? **Veterinary Microbiology**, v. 165, n. 1–2, p. 123–134, 2013.

SCICLUNA, M. T. et al. Validation of an indirect ELISA employing a chimeric recombinant gag and env peptide for the serological diagnosis of equine infectious anemia. **Journal of Virological Methods**, v. 251, n. September 2017, p. 111–117, 2018.

SILVA, R. A. M. S. et al. Equine viral diseases in the Pantanal, Brazil: studies carried out from 1990 to 1995. **Revue d'Élevage et de Médecine Vétérinaire des Pays Tropicaux**, v. 52, p. 9-12, 1999a.

SILVA, R. A. M. S. et al. Swamp Fever in wild horses from the Pantanal, Brazil. **Revue D'Élevage et Médecine Vétérinaire des Pays Tropicaux**. v.52, p.99-101, 1999b.

SILVA, R. A. M. S.; ABREU, U. G. P. de; BARROS, A. T. M. de. Anemia Infeciosa

Eqüina : Epizootiologia , Prevenção e Controle no Pantanal. **Embrapa: Pantanal - Circular Técnica**, 29, p. 1–32, 2001.

SÖDING, J.; BIEGERT, A.; LUPAS, A. N. The HHpred interactive server for protein homology detection and structure prediction. **Nucleic Acids Research**, v. 33, n. SUPPL. 2, p. 244–248, 2005.

SOUTULLO, A. et al. Systematic epitope analysis of the p26 EIAV core protein. **Journal of molecular recognition : JMR**, v. 20, p. 227–237, 2007.

SOUZA, G. da S. e et al. Factors affecting the performance of Pantaneiro horses. **Revista Brasileira de Zootecnia**, v. 47, 2018.

ULLMANN, L. S. et al. Comparative clinical sample preparation of DNA and RNA viral nucleic acids for a commercial deep sequencing system (Illumina MiSeq®). **Journal of virological methods**, v. 220, p. 60–3, ago. 2015.

ZHENG, Y. H. et al. In vivo dynamics of equine infectious anemia viruses emerging during febrile episodes: insertions/duplications at the principal neutralizing domain. **J Virol**, v. 71, n. 7, p. 5031–5039, 1997.

5 Artigo Científico 2

A quantitative PCR to *Equine infectious anemia virus* from Brazil: evidences of serological silent equids require a new diagnosis workflow to EIA control

Camila Dantas Malossi¹, Márcia Furlan Nogueira Tavares de Lima², Fábio Sossai Possebon¹, Daniel de Moura Aguiar⁴, Erna Greessien Kroon³, Leila Sabrina Ullmann¹, João Pessoa Araujo Junior^{1*}.

1. São Paulo State University (Unesp), Institute of Biotechnology (IBTEC), Botucatu, SP, Brazil. Address: Alameda das Tecomarias, s/n, Chácara Capão Bonito, CEP: 18607-440, Botucatu, SP, Brazil. Phone No.: +55 14 3880-0423.
2. EMBRAPA Pantanal, Corumbá, MS, Brazil.
3. Virology Laboratory, Minas Gerais Federal University, Belo Horizonte, MG, Brazil.
4. Virology and Rickettsiosis Laboratory, Mato Grosso Federal University, Cuiabá, MT, Brazil.

*Corresponding Author: João Pessoa Araujo Junior. Address: Alameda das Tecomarias, s/n, Chácara Capão Bonito, CEP: 18607-440, Botucatu, SP, Brazil, CEP: 18607-440. Phone No.: + 55 14 3880-0423 e-mail: joao.pessoa@unesp.br

Abstract

Equine infectious anemia virus (EIAV) is a lentivirus that causes equine infectious anemia (EIA), a persistent viral infection. The recommended diagnosis method is the gold standard agar gel immunodiffusion (AGID). Some countries use enzyme-linked immunosorbent assay (ELISA) and immunoblot as complement. Although OIE suggests a PCR to conflicting results of serological tests, the primers employed do not amplify the virus currently circulating in Brazil due to the higher genetic variation. A particular case in Brazil is Pantanal areas, where the disease is endemic and euthanasia is not mandatory. In order to find a more sensitive and specific molecular technique to detect the viral Brazilian strain, we designed a quantitative PCR based on EIAV sequences from Pantanal – Brazil to detect the current virus. Two hundred fifty five equid samples were tested in AGID and ELISA and their gDNA extracted were tested with qPCR and a previously described Semi Nested PCR. In total, 179 equids were positive in any test, with 98 positives by Semi Nested PCR and 122 positives by qPCR, the same positive number for AGID. We observed conflitant results for each method, with the same equid positive for AGID, ELISA or qPCR and negative for the others. We propose a diagnostic sequence to minimize the presence of false negatives in the Brazilian Pantanal and help to control the disease in this endemic area.

Keywords: Equine infectious anemia virus, qPCR, Brazil, LTR variation, Equids, Diagnosis

Background

Equine infectious anemia virus (EIAV) is a lentivirus that causes equine infectious anemia (EIA), a persistent viral infection with almost worldwide distribution (ISSEL et al., 2014). This disease causes fever and thrombocytopenia in an acute phase and, episodic bouts of fever, thrombocytopenia, anemia and depressed neurologic signs in a chronic phase. If the horse survives, the frequency of disease episodes will gradually decrease and it will progress to an inapparent carrier state, where no clear clinical signs are observed (CAPPELLI et al., 2017). In while world, EIA has mandatory notification and positive animals euthanasia. In Brazil, EIAV is endemic in Pantanal region, and differently from other regions of Brazil, euthanasia is not mandatory for positive equids. Also, in these locals, there are projects to separate and control positive farms from negative ones (NOGUEIRA et al., 2017).

Diagnosis is the principal key to EIA control: false negative equids allow new infections and keep the disease circulating. World Organization for Animal Health (OIE) recommends serological tests to detect serum antibody to EIAV and to confirm the diagnosis of EIAV infection. Agar gel immunodiffusion (AGID) test (COGGINS; NORCROSS; NUSBAUM, 1972) is the gold standard test and enzyme-linked immunosorbent assay (ELISA) (Suzuki et al., 1982) should be used to confirm the AGID diagnosis. The results can also be confirmed by the immunoblot technique. EIAV diagnosis based on its in vitro isolation is complicated as it does not easily replicate in cell cultures (HINES; MAURY, 2001). A polymerase chain reaction (PCR) assay may detect EIAV proviral DNA from the peripheral blood of horses and OIE recommends nested PCR described by NAGARAJAN and SIMARD (2001) or the real-time reverse-transcriptase PCR assay by COOK and collaborators (2002). Molecular diagnosis maybe used for the detection of EIAV infection in horses only in specific situations: i) Conflicting results on serologic tests; ii) Suspected infection but negative or questionable serologic results; iii) Complementary test to serology for the confirmation of positive results; iv) Confirmation of early infection, before serum antibodies to EIAV develop; v) Ensuring that horses that are to be used for antiserum or vaccine production or as blood donors are free of EIAV; vi) Confirmation of the status of a foal from an infected mare (OIE, 2013). Even molecular diagnosis may have a low sensitivity as it is influenced by the genetic heterogeneity of the virus and the temporal variation of viral loads during infection (LEROUX; CADORÉ; MONTELARO, 2004; SCICLUNA et al., 2013).

AGID test become internationally recognized in 1976 after its use EIA control in USA, when the positive correlation of this test and horse inoculation test was confirmed (ISSEL et al., 2013). Ever since, this test has been the only one accepted as an official form of diagnosis in many countries, with few using complementary serological methods (ELISA and Western blot) or direct methods (PCR) (COOK; LEROUX; ISSEL, 2013; RICOTTI et al., 2016). However, many countries, such Italy and Argentina, already report the presence of false negative animals for AGID (ISSEL et al., 2013; RICOTTI et al., 2016) and question the exclusive use of this test for diagnosis. Italy adopts 'three-tiered' diagnostic system, using ELISA tests first, followed by AGIDT, and immunoblot tests where indicated (ISSEL et al., 2013).

A reverse transcriptase nested PCR (RT-nPCR) assay for the detection of viral gag gene sequences in plasma from EIAV-infected field horses has been described (LANGEMEIER et al., 1996) and a nested PCR assay for proviral DNA detection using the same gene too (NAGARAJAN; SIMARD, 2001). Another PCR and RT-PCR protocols had already tested for EIAV diagnosis (CAPOMACCIO et al., 2012; CAPPELLI et al., 2011; COOK et al., 2002; KUHAR et al., 2014; QUINLIVAN; COOK; CULLINANE, 2007). Although these techniques showed greater sensitivity early in infection than the agar gel immunodiffusion test, plasma viral RNA (viremia) may not be detected in all horses with subclinical infections (CAPPELLI et al., 2011; ISSEL et al., 2013). The Nagarajan and Simard nested PCR do not amplified field viruses sequences (CAPPELLI et al., 2011). Modifications in primers from DONG and collaborators (2012) reaction can amplify Brazilian EIAV (CURSINO et al., 2018), but this test is not used in EIA diagnostic routine in Brazil.

In this study, we develop of an EIAV-specific real-time PCR (quantitative PCR or qPCR) assay, using primers in the tat region for the direct detection of EIA proviral DNA from peripheral blood leukocytes of naturally infected field equids from Brazil and the sensitivity of the assay is compared with semi-nested PCR (DONG et al., 2012), rgp90 ELISA (REIS et al., 2012) and serological Coggins test (AGID). The presented qPCR were tested in 255 Pantanal equids from controlled farms (negative animals) and infected farms (positive animals).

Material and Methods

1. Sample collection

Serum and leukocyte layer were collected from 255 horses and mules from Pantanal, Brazil (Figure 1) since 2014 until 2017. They may present EIA symptoms or not. These equids were accompanied by EMBRAPA project for EIA control, that supervised positive and negative farms at Pantanal region. Samples were refrigerated at 4°C and than frozen at -20°C until used for nucleic acid extraction or serological tests.

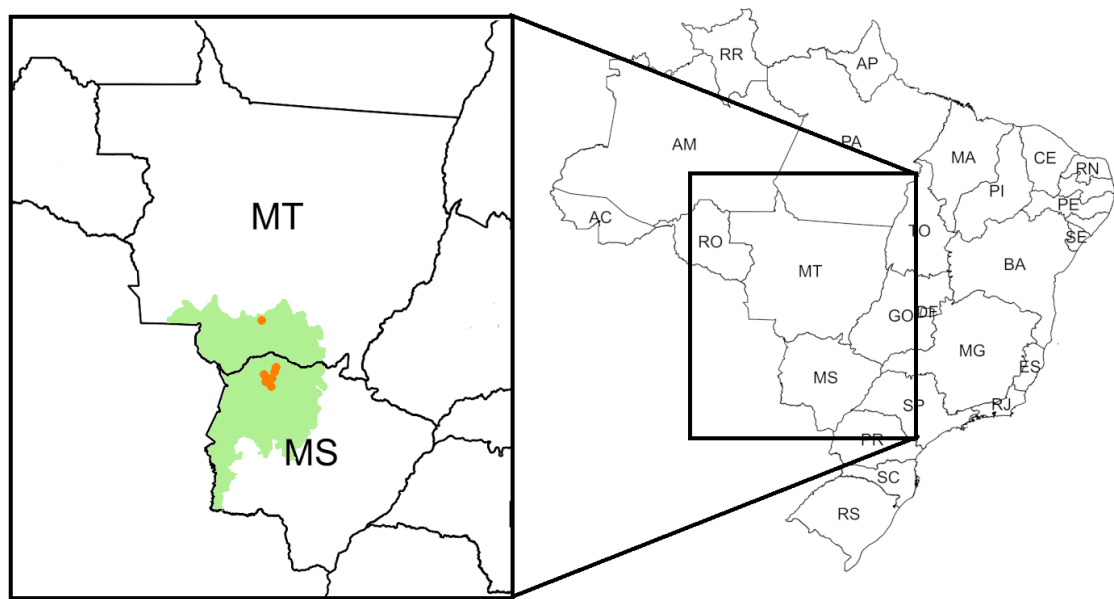


Figura 1: The map shows the location of the farms from which the samples were collected. Pantanal location is showed in green, farms in orange.

2. Positive and negative controls

The positive sample used is positive for EIAV according to AGID, rgp90-ELISA, SemiNested and qPCR. This plasma sample (BRA1) had the RNA extracted previously and the genome completely sequenced (MALOSSI et al., 2019 – not published). The negative sample is originated from a controlled farm, in which there are no cases of EIA. This sample was negative according to AGID, rgp90-ELISA, SemiNested PCR e qPCR.

3. Serological tests

Equids serum were tested with agar gel immunodiffusion (AGID) and enzyme-linked immunosorbent assay (ELISA) in Embrapa Pantanal laboratory. AGID utilize the p26 antigen of EIAV, like described in COGGINS; NORCROSS; NUSBAUM (1972). The ELISA used in this study targets the recombinant gp90, based on a brazilian sequence (REIS et al., 2012).

4. DNA purification

Total equid genomic DNA was extracted from 100 uL of total blood using a Illustra Blood GenomicPrep Mini Spin Kit (GE Healthcare Life Sciences) according to the manufacturer's instructions. Quality and quantity assessments of each DNA preparation were carried out using a NanoDrop 1000 spectrophotometer.

5. SemiNested PCR analysis

Semi Nested PCR used to evaluate this samples is based on that described by Dong and collaborators (2012), amplifying a fragment of LTR U5 region and tat gene of the EIAV genome. Only three primers of four described was used; EIAVltr-1F do not anealle in EIAV current in Brazil (CURSINO et al., 2018). The Semi Nested PCR consists of 20uL solution containing 2X GoTaq Green Master Mix (Promega), 0,3 mM of each primer, sample and nuclease-free water to complete the reaction. Two microliters of the equine total DNA was used for the first reaction, and we used 1 uL of the amplified DNA product of the first reaction (PCR) in second reaction. The conditions of the first PCR cycles consisted to a start denaturation for 10 minutes at 95° C, and 35 repeated cycles of 30 seconds at 95°C, 30 seconds at 56°C, and 1 minute at 72°C, followed by a 10 minutes final extension at 72°C. At Semi Nested PCR, the initial denaturation was at 95°C for 5 minutes, and 30 repeat cycles of 30 seconds at 95°C, 30 seconds at 58°C, and 1 minute at 72°C, followed by a 10 minutes final extension at 72°C. The positive and negative controls were included in each reaction, plus the reaction control (water). The PCR and Semi Nested PCR products were submitted to a horizontal electrophoresis in 1,5% agarose gel, plus 1uL of SYBR Safe DNA gel Stain (ThermoScientific) for each 10 mL of gel, with buffer TBE 1X pH 8,0 (44,58 M Tris-base; 0,44 M boric acid; 12,49 mM EDTA) at 100V /50mA, and visualized under ultraviolet light (UV).

6. Primers design and quantitative PCR

Quantitative PCR (qPCR) primer forward was designed with Primer3 software (ROZEN; SKALETSKY, 2000) using genome sequence of Brazilian EIAV (BRA1 and BRA2) to amplify a 71 bp product together the Dong PCR reverse primer (Table 1). Primer was analyzed by Oligo Analyzer from Integrated DNA Technologies (<https://www.idtdna.com/calc/analyzer>) and reviewed for match sequences in NCBI BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). GoTaq® qPCR Master Mix (Promega) was used in the reaction plus of each primer, 4 uL of gDNA sample (about 100 ng) and nuclease free water to 20 uL. The thermocycling program set up in a 7500 Fast qPCR (Applied Biosystems) was 1 cycle at 95°C for 10 min, 40 cycles at 95°C for 15 s and 60 °C for 1 min, followed by the standard melting curve. Equine G3PDH (GARTON et al., 2002) was amplified in each sample comproved the DNA purification.

The positive control sample was tested in a serial dilution in nuclease-free water with 1:10 dilution factor to compare the analytical sensitivity of the proposed qPCR reactions with the previously used Semi Nested PCR. Amplification curve of qPCR were compared with the PCR and Semi Nested PCR products analized in a horizontal electrophoresis in 1,5% agarose gel.

Table 1. Tested primers to SemiNested PCR and qPCR.

	Primer	Sequence	Amplicon length	Reference
qPCR	EIAV F	GAA CCT GGC TGA TCG TAG GA	71 bp	this work
	EIAV ltr-1R	CAG GAA CAC CTC CAG AAG AC		Dong (2012)
Semi Nested PCR	EIAV ltr-2F	ATT CTG CGG TCT GAG TCC	229 bp	Dong (2012)
	EIAV ltr-1R	CAG GAA CAC CTC CAG AAG AC		Dong (2012)
	EIAV ltr-2F	ATT CTG CGG TCT GAG TCC CT	198-203 bp	Dong (2012)
	EIAV ltr-2R	TAA GTT CTC CTC TGC TGT		Dong (2012)

7. Sanger sequencing

Sanger sequencing was performed with BigDye Terminator v3.1 Cycle Sequencing kit (Life Technologies) in two reactions following manufacturer's protocol, one with forward primer, other with reverse one. Capillary electrophoresis was performed in an ABI 3500 automatic sequencing system (Life Technologies) and analyzed with MEGA 7 software (KUMAR; STECHER; TAMURA, 2016).

8. Statistical analysis

Sensitivity was calculated as true positive/(true positive + false negative) whereas specificity was true negative/(true negative + false positive). The results of calculations were expressed as percentages. The Statistical data analysis was realized by SAS (Statistical Analysis System). Qui-Squared and Kappa were used to verify the concordance between the qPCR and Semi Nested PCR results, both with 95% of confidence interval (CI95).

9. Flowchart sequence

Based on the results from 255 animals, it was analyzed which sequence would allow a greater detection of positive animals and decrease the chance of a false negative. For each sequence, it was analyzed the percentage of positive detection of the diagnostic flowchart, the prevalence of positives according to that flowchart, and 95% CI.

Results

1. qPCR standardization

The couple of primers tested for qPCR results in a clean reaction, without nonspecific reactions, and with a single peak melting curve with approximate T_m of 79°C. The qPCR amplified product was sequenced and the fragment obtained present a 100% identity with the sequence obtained on the massive sequencing, for the same sample. All analyzed samples amplified the same viral target.

In the analytical sensitivity test, the two reactions, qPCR and Semi Nested PCR, amplified up to the same dilution factor, demonstrating the same sensitivity (Figure 2). The analysis of the tat region in which the primers annealed, shown that there are alterations on the 3' region of the primer forward that may result in an amplification difficulty of the Miyazaki strain and the culture virus DLV9-7 (Figure 3). A degenerate primer for those regions was analyzed but presented a delta G near to -9 kcal.mol⁻¹, which led to proceed with the primer without degeneration.

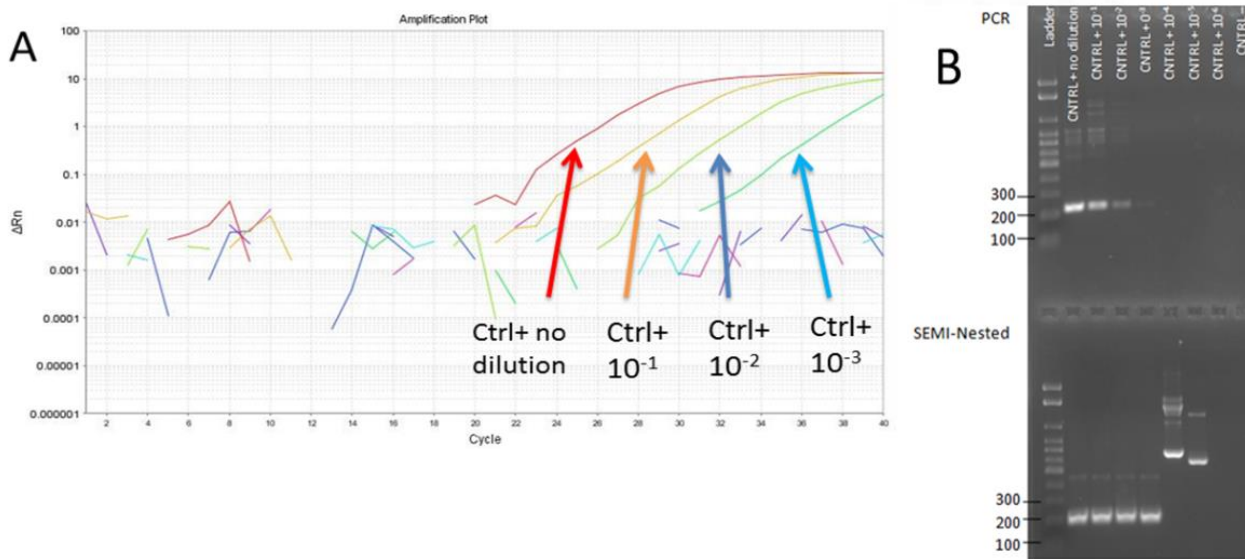


Figure 2. Comparison of the qPCR and Semi Nested PCR sensibility. (A) The amplification curve shows that qPCR amplifies to dilution 1:1000 and (B) the electrophoresis gel shows that Semi Nested PCR amplifies to the same dilution.

2. LTR sequences show nucleotides variations

The Semi Nested PCR product allows the sequencing of approximately 200 bp of the viral LTR. Thirty six samples amplified by Semi Nested PCR were sequenced to sequence confirmation, and two of them (5,55%) presented sequences that do not correspond with EIAV: one corresponding to an *Equus caballus* microsatellite region (GenBank Acession FN419718.1) and the other to an *Equus asinus* cytochrome c oxidase assembly protein COX19 (GenBank Acession XR_001398933.1).

Accession	Sequence
Wyoming_AF033820_Tat	CTGTTGAACCTGGCTGATCGTAGGA
Miyazaki_JX003263_Tat	...C...T...A...TA...
ITA_SA_KM247555_Tat	...C...TA...
ITA_DE_KM247554_Tat	...C...TA...
IRE_H3_JX480634_Tat	...C...TA...
IRE_F4_JX480633_Tat	...C...TA...
IRE_F3_JX480632_Tat	...C...TA...
IRE_F2_JX480631_Tat	...C...TA...
DV10-3_HM141910_Tat	...C...TA...
DLV9-7_HM141922_Tat	...C...A...TA...
BRA2_Tat	...C...TA...
BRA1_Tat	...C...TA...

Wyoming_AF033820_Tat	ACAGAA	GTCTTCTGGAGGTGTTTCCTG	GCCAGAACACAGGAGGACAGCAGG
Miyazaki_JX003263_Tat	C.....C.....	TTTC	
ITA_SA_KM247555_Tat	C.....	G.....A.....G.AC	
ITA_DE_KM247554_Tat	C.....	G.....A.....G.AC	
IRE_H3_JX480634_Tat	C.....	G.....A.....G.AC	
IRE_F4_JX480633_Tat	C.....	G.....A.....G.AC	
IRE_F3_JX480632_Tat	C.....	G.....A.....G.AC	
IRE_F2_JX480631_Tat	C.....	G.....A.....G.AC	
DV10-3_HM141910_Tat	C.....	A.....G.AT	
DLV9-7_HM141922_Tat	C.....	A.....G.AT	
BRA2_Tat	C.....	G.....GA.....G.A.	
BRA1_Tat	G.....	G.....A.....G.A.	
Wyoming_AF033820_Tat	AAGCAAGACCCAACTACCATTGTCAGCTGTGTTTCCTGAGGTCTCTAGGA		
Miyazaki_JX003263_Tat	.G.ATTTC..GG....T.....C.....C.....G...		
ITA_SA_KM247555_Tat	CTC..CC.GT..GT.....A.....C...T...A...T.G...		
ITA_DE_KM247554_Tat	CTC..CC.GT..GT.....A.....C...T...A...T.G...		
IRE_H3_JX480634_Tat	CTC..CC.GT..GT.....A.....C...T...A...T.G...		
IRE_F4_JX480633_Tat	CTC..CC.GT..GT.....A.....C...T...A...T.G...		
IRE_F3_JX480632_Tat	CTC..CC.GT..GT.....A.....C...T...A...T.G...		
IRE_F2_JX480631_Tat	CTC..CC.GT..GT.....A.....C...T...A...T.G...		
DV10-3_HM141910_Tat	C.C..CC.GT..G...T.....A.....A...AT.G...		
DLV9-7_HM141922_Tat	C.C..CC.GT..G...T.....A.....A...AT.G...		
BRA2_Tat	.GC..TC.GT.....T.....A.....T.G...G		
BRA1_Tat	.GC..CC.GT.....T.....CT.G...G		
Wyoming_AF033820_Tat	ATTGATTACCTCGATGCTTCATTAAGGAAGAAGAATAAACAAAGACTGAA		
Miyazaki_JX003263_Tat	T...T..CCAC..CC.G.A.....A.....G...G.GGCA..		
ITA_SA_KM247555_Tat	C...T.G...C.C...C.G.A.....C.....GAA...		
ITA_DE_KM247554_Tat	C...T.G...C.C...C.G.A.....C.....GAA...		
IRE_H3_JX480634_Tat	C...T.G...C.C...C.G.A.....C.....GAA...		
IRE_F4_JX480633_Tat	C...T.G...C.C...C.G.A.....C.....GAA...		
IRE_F3_JX480632_Tat	C...T.G...C.C...C.G.A.....C.....GAA...		
IRE_F2_JX480631_Tat	C...T.G...C.C...C.G.A.....C.....GAA...		
DV10-3_HM141910_Tat	C.....T..CAGC..GC.G.AA.....C.....AA..		
DLV9-7_HM141922_Tat	C.....T..CAGC..GC.G.A.....C.....A...		
BRA2_Tat	C.....T..CT.....A.....G...G.A...		
BRA1_Tat	C.....T..CA.....G.A.....G...G.AA..		
Wyoming_AF033820_Tat	GGCAATCCAACAAGGAAGACAACCTCAATATTTGTTATAA		
Miyazaki_JX003263_Tat	..A...A..G.GT...C...G...G...C.....		
ITA_SA_KM247555_Tat	.CA.C.G..G.G...GA...G...A...CC..C.....		
ITA_DE_KM247554_Tat	.CA.C.G..G.G...GA...G...A...CC..C.....		
IRE_H3_JX480634_Tat	.CA.C.G..G.G...GA...G...A...CC..C.....		
IRE_F4_JX480633_Tat	.CA.C.G..G.G...GA...G...A...CC..C.....		
IRE_F3_JX480632_Tat	.CA.C.G..G.G...GA...G...A...CC..C.....		
IRE_F2_JX480631_Tat	.CA.C.G..G.G...GA...G...A...CC..C.....		
DV10-3_HM141910_Tat	...C...A.GG.G.....C.....G...C.....		
DLV9-7_HM141922_Tat	...C...A.GG.G.....C.....G...C.....		
BRA2_Tat	...T...AGGG.G.....C.....G...C.....		
BRA1_Tat	...T...T...G...G.....G...C.....		
	..AG..T.T..GG..G.....G...C.....G...		

Figura 3: Alignment of *tat* gene from EIAV Brazilian sequences (BRA1 and BRA2), EIAV Wyoming (AF033820), EIAV Miyazaki (JX003263), EIAV Italy (KM247554 and KM247555), EIAV Ireland (JX480631-JX480634), EIAV DV10-3 (HM141910) and DV9-7 (HM141922). Annealing of qPCR primers are shown in yellow.

3. qPCR presents a higher EIAV detection than SemiNested PCR.

From 255 equids (horses and mules) analyzed for the presence of EIAV infection, 98 were positive by Semi Nested PCR and 122 were positive by qPCR (Table 2). From the positive animals in qPCR, 105 were positive in IDGA or ELISA and from the negative animals, 57 presented positivity in AGID or ELISA. In Semi Nested PCR, 87 positive animals were also positive in serological tests and 75 of the negative animals were positive in AGID/ELISA. In total, 162 animals were positive in any serological tests, AGID or ELISA; 179 animals were positive in any test (serological or molecular). Of the four methods tested, only Semi Nested PCR did not present any positive animals with this technique alone (Figure 4).

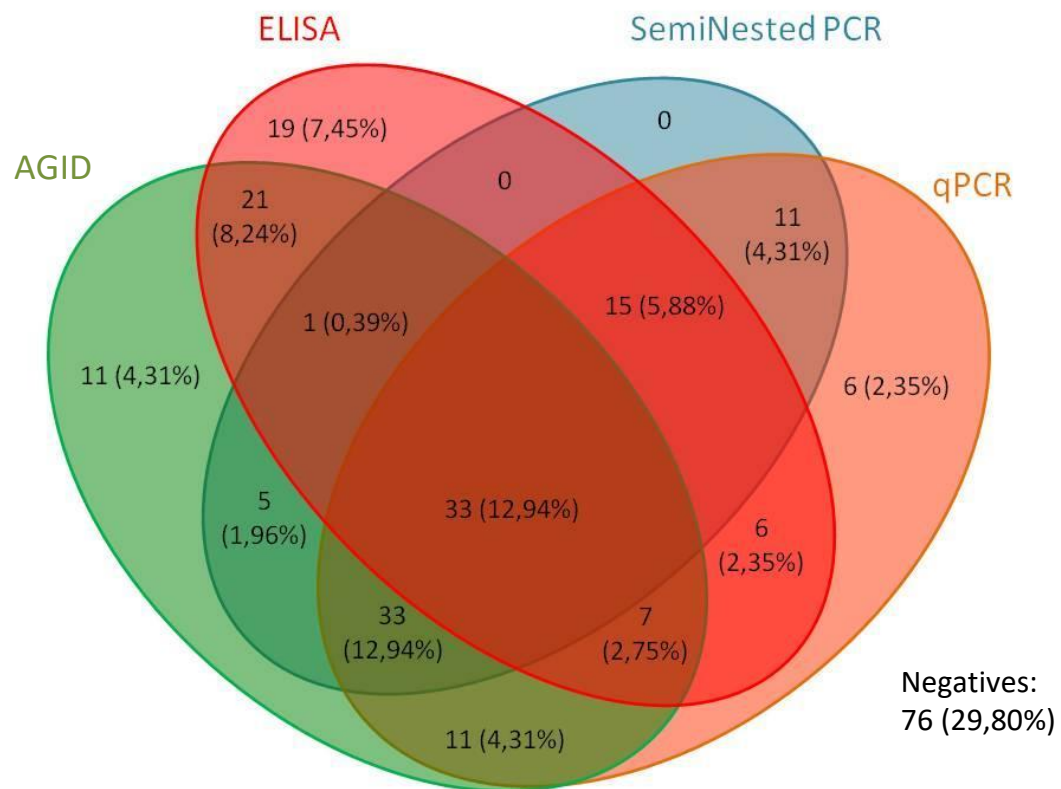


Figure 4: Representative diagram of positive samples in each method: AGID, ELISA, Semi Nested PCR and qPCR.

Considering the total number of positive as positive samples in at least one test, we can establish a detection rate of 68% (122/179) for AGID, 57% (102/179) for the ELISA, 68% (122/179) for qPCR and 55% (98/179) for the Semi Nested PCR. Thus, the detection of qPCR is equal to the OIE gold standard (AGID), although the animals detected were not the same (Figure 4).

Table 2. Table 2 x 2 with comparison between qPCR and another tested methods: (A) Semi Nested PCR, (B) AGID e (C) ELISA.

A

		qPCR		
		Negative	Positive	Total
Semi Nested PCR	Negative	125 (49%)	30 (11.7%)	155 (60.7%)
	Positive	9 (3.5%)	91 (35.6%)	100 (39.2%)
Total		134 (52.5%)	121 (47.4%)	255

B

		qPCR		
		Negative	Positive	Total
AGID	Negative	95 (37.3%)	38 (14.9%)	133 (52.2%)
	Positive	38 (14.9%)	84 (32.9%)	122 (47.8%)
Total		133 (52.2%)	122 (47.8%)	255

C

		qPCR		
		Negative	Positive	Total
rgp90 - ELISA	Negative	92 (36.1%)	61 (23.9%)	153 (60%)
	Positive	41 (16.1%)	61 (23.9%)	102 (40%)
Total		133 (52.2%)	122 (47.8%)	255

4. Diagnostic tests comparison

The qPCR test, using AGID as gold standard reference, presents relative diagnostic sensitivity and relative diagnostic specificity of 71.4 % and 68.9 %, respectively. The accuracy was calculated in 70.2 %, with 95% CI between 65.4 and 76.8 %.

Compared to Semi Nested PCR, qPCR is a more sensitive and specific test. The concordance between qPCR and Semi Nested PCR was 91.3%, with a Kappa index of 0.78. By the chi-square test, p was lower than 0.0001. The qPCR still detected 30 other animals that Semi Nested had not detected, 18 positive for EIAV by IDGA and 6 positive for ELISA, suggesting a better detection of positive animals through this molecular test.

5. EIAV diagnosis flowchart

A diagnosis serial flowchart is logical and cost-effective, with the first tests being always the most sensitive and the second test the most specific, to eliminate false-positives. We evaluate diverse diagnosis flowcharts using the data obtained on the four tests to eliminate as much as possible the need to check for false negatives in an endemic area such as the Pantanal. The Semi Nested PCR did not detect alone any animal that has not been detected for another test, being thus removed from the flowcharts. We took into consideration the creation of flowcharts with at least on a serological test. Using the data from this study, the diagnosis sequence that is capable of detect 100% of the animals with less cost is AGID followed by ELISA and ending with qPCR (Table 3). The tests were designed so as to follow the still recommendation of AGID as the gold standard, however including more sensitive exams and with different targets and methodologies.

Table 3: Test of diagnosis flowchart for EIAV.

Diagnosis sequence	n positives in some test	% of positive samples in some test	Prevalence	IC 95%
AGID-ELISA	162	91%	63.53	57,30-69,45
AGID-qPCR	160	89%	62.75	56,49-68,70
AGID-ELISA-qPCR	179	100%	70.20	64,17-75,74
ELISA-qPCR	163	91%	63.92	57,70-69,82

Discussion

After the description and validation of the AGID to EIA, its use was immediately adopted by the international veterinary community. Despite the standardized ELISA tests and

with good results in the diagnosis of EIA, the AGID continue to be the gold standard serological test due to its proven correlation with results in test of horse inoculation for EIAV and high precision (ISSEL et al., 2013). However, Brazil presents a endemic situation in Pantanal, where there is a high prevalence of EIA and non-euthanized animals, but segregated in controlled farms (NOGUEIRA et al., 2017). The presence of false negative animals makes it very difficult to control the disease in the region, either on farms known to be negative, or on animals released for transport.

In this paper 57 of 255 (22.35 %) studied animals were negative in AGID and positive in other tests, a concerning number that keeps EIA circulating and contaminating previously negative animals. Thus, the validation of new diagnostic methods to reduce the number of false negatives is imperative to aid in the control of EIA in Pantanal. Researchers around the world have already questioned the only use of AGID as a diagnostic test (COOK; LEROUX; ISSEL, 2013; ISSEL et al., 2013; RICOTTI et al., 2016).

AGID and ELISA have already been compared for sensitivity and false-negatives of AGID previously (ISSEL et al., 2013; RICOTTI et al., 2016). The ELISAs used for this comparison was a test based on a recombinant p26, same target recognized by AGID. Although the ELISA presents a better sensitivity than AGID for serological detection, the results must be cautiously observed due to the presence of false negatives. In a control setting where the animal will not be euthanized (Pantanal) the epidemiologic risk of a false positive is less important than a false negative. Since AGID has a lower sensitivity than ELISA's tests, these have been standardized to detect animals with low antibody levels (ISSEL et al., 2013). A current example is a recently released commercial kit in Italy that uses a chimeric protein recombinant of gag and env as a target (SCICLUNA et al., 2018). The ELISA used in this study targets the recombinant gp90 based on a Brazilian sequence, different from AGID's p26 (REIS et al., 2012). The rgp90-ELISA has an agreement of 95.42% with agar gel immunodiffusion test results (REIS et al., 2012). The serological detection using different targets is interesting because, if the false negative is caused by a viral variant non-reactive to that target, antibodies for another protein may be detected. However, the envelope glycoprotein gp90 shows greater variation among all genes translated from the virus, which may also present, in an endemic situation like the Pantanal, false-negative animals. In Italy, once presented with contradictory results from AGID and ELISA, it realizes an immunoblot for three targets (p26, gp45, and gp90) to confirm the infection (ISSEL et al., 2013). Besides

the three tests substantially improve the diagnostic accuracy when used together, all of them are serological exams and will present the same difficulties, for example, in foals that receive maternal antibodies through breastfeeding or animals with recent infection

Polymerase Chain Reaction (PCR) is a direct diagnosis test that detects the genetic material from the tested pathogen. Detecting the proviral DNA offers a more accurate diagnostic in multiple disease stages (chronic, acute and inapparent), since the use of genomic RNA by RT-qPCR would only allow the detection of animals in viremia (CAPPELLI et al., 2011). Although not used in this study, one step RT-qPCR with RNA and DNA samples is possible, in order to increase the sensitivity of the test by combining detection of the viral and proviral genomes. Kits that perform the simultaneous extraction of the two materials are already commercial available for such methodology.

In association with this, a study about the gene variation must be conducted to choose the most conserved region possible. The *tat* gene was chosen to be amplified in a region conserved in most lineages, in which only the Miyazaki lineage would not have correspondence (no qPCR test with this lineage). Dong and collaborators (2012) already used this region jointly with the U5 region of LTR for viral detection and one of their described primers was used in our reaction.

The Semi Nested PCR derivative of primers initially described by Dong and collaborators (2012) was standardized as a possible technique to amplify a fragment of the proviral genome of the circulating AIEV in the Pantanal. The high risk of contamination of a Semi Nested PCR and the practicality and sensitivity offered by quantitative PCR makes this second a better option in terms of molecular diagnosis. Quantitative PCR showed the same analytical sensitivity as the Semi Nested PCR and detected more positive samples (confirmed by AGID). When compared to the AGID (gold standard), qPCR detected the same number of positive animals as the serological test, however in different animals (Figure 4). This demonstrates that none of the tests is detecting 100% of the positive animals in the Pantanal.

Faced with this situation of diagnostic tests with different sensitivities and specificities, we propose a diagnostic sequence to minimize the presence of false negatives in the Brazilian Pantanal. This sequence was designed so that the initial test was the one recommended by the MAPA, besides being the least expensive test in the market. Positive samples for the gold standard test are retested to confirm the diagnosis. Negative samples are tested in the other two methods for confirmation. Positive result in any of the serological test

is reported as positive samples, and positive samples only in qPCR are reported as positive requiring monitoring. The ideal situation to confirm positive cases only in qPCR should be the sequencing of the amplified product. With this diagnostic sequence, a total of 100% of the animals diagnosed as positive in this study would be detected (Table 3). As every animal detected by the Semi Nested PCR were also detected by another method, this test was withdrawn from the diagnostic sequence. With this suggested flowchart, there is a chance to get false positives.

The euthanasia of positive animals is mandatory for other areas of Brazil where the EIA is not endemic as Pantanal. In this case, the occurrence of false positives is not desirable. Therefore, in cases like this, we suggest that animals are initially tested using both serological tests, AGID and ELISA. The animal is only considered negative with the two confirmatory results. Conflicting results should be forwarded to the qPCR test (Figure 6). Positive results only in the ELISA can be considered as false positives. However, this study observed 7.45% (19/255) of positive animals only in this test, a value much higher than the 3.6% false positive rate calculated in rgp90-ELISA (REIS et al., 2012). As we do not know what places in Brazil the animals tested in the description of this ELISA are, there is the possibility of the high genetic variation present in an endemic environment such as the Pantanal change these values. Thus, positive results only in the ELISA should be considered with caution in non-endemic regions (Figure 6).

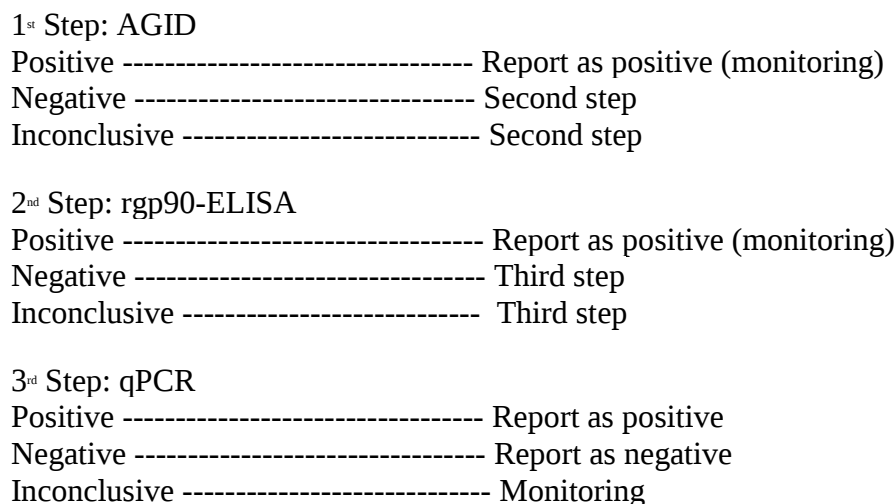


Figure 5. Diagnostic algorithm proposed to minimize the presence of false negatives in the Brazilian Pantanal. Samples will be tested for AGID, ELISA, and qPCR, respectively, and a positive for a serological test already qualifies the animal as positive, and the positive only in

the direct examination requires a follow-up of the animal and repetition of the tests posteriorly (6 months).

1st Step: AGID and rgp90-ELISA

Positive in both ----- Report as positive
 Positive only in AGID ----- Second step
 Positive only in ELISA ----- Second step
 Negative in both ----- Negative
 Inconclusive in some test ----- Second step

2nd Step: qPCR

AGID Positive / qPCR Positive ----- Report as positive
 AGID Positive / qPCR Negative ----- Retest with another sample

Retest: AGID + / qPCR +: Positive
 Retest: AGID + / qPCR - : Positive
 Retest: AGID - / qPCR +: Positive
 Retest: AGID - / qPCR - : Negative

ELISA Positive / qPCR Positive ----- Report as positive

ELISA Positive / qPCR Negative ----- Retest with another sample

Retest: ELISA + / qPCR +: Positive
 Retest: ELISA + / qPCR - : Negative (monitoring)
 Retest: ELISA - / qPCR +: Positive
 Retest: ELISA - / qPCR - : Negative

Inconclusive in qPCR ----- Consider the result of AGID (with two retests)

Figure 6. Diagnostic algorithm proposed to non endemic regions in Brazil. Samples will be tested for AGID and ELISA, if there is conflitant results, qPCR. Conflitant results with qPCR need to be monitored for possibility of false positive. A retest with another sample is recommended and monitoring during the property quarantine.

Conclusions

Our data indicate that a significant number of equines infected by EIAV can escape detection if only the AGID test is performed. To actually implement an effective control plan and minimize the chance of keeping positive animals in the middle of a controlled farm, we suggest implementing other available tests in a diagnostic sequence. Currently, many Brazilian laboratories have the necessary structure to perform such exams. The algorithm was designed in order to prioritize the already well established and low-cost test, so that an initial screening is performed and it is not necessary for all samples to pass all the tests. In spite of the analyzes, none of the tests presented total efficiency of detection of the infection, making

the search for new diagnostic techniques for EIAV essential for the control of the disease in Brazil.

References

- CAPOMACCIO, S.; WILLAND, Z. A.; COOK, S. J.; ISSEL, C. J.; SANTOS, E. M.; REIS, J. K. P.; COOK, R. F. Detection, molecular characterization and phylogenetic analysis of full-length equine infectious anemia (EIAV) gag genes isolated from Shackleford Banks wild horses. **Veterinary Microbiology**, v. 157, n. 3–4, p. 320–332, 2012.
- CAPPELLI, K.; CAPOMACCIO, S.; COOK, F. R.; FELICETTI, M.; MARENZONI, M. L.; COPPOLA, G.; VERINI-SUPPLIZI, A.; COLETTI, M.; PASSAMONTI, F. Molecular detection, epidemiology, and genetic characterization of novel european field isolates of equine infectious anemia virus. **Journal of Clinical Microbiology**, v. 49, n. 1, p. 27–33, 2011.
- CAPPELLI, K.; COOK, R. F.; STEFANETTI, V.; PASSAMONTI, F.; AUTORINO, G. L.; SCICLUNA, M. T.; COLETTI, M.; VERINI SUPPLIZI, A.; CAPOMACCIO, S. Deep sequencing and variant analysis of an Italian pathogenic field strain of equine infectious anaemia virus. **Transboundary and Emerging Diseases**, v. 64, n. 6, p. 2104–2112, 2017.
- COGGINS, L.; NORCROSS, N. L.; NUSBAUM, S. R. Diagnosis of equine infectious anemia by immunodiffusion test. **American Journal of Veterinary Research**, v. 33, n. 1, p. 11–18, 1972.
- COOK, R. F.; COOK, S. J.; LI, F.; MONTELARO, R. C.; ISSEL, C. J. Development of a multiplex real-time reverse transcriptase-polymerase chain reaction for equine infectious anemia virus (EIAV). **Journal of Virological Methods**, v. 105, n. 1, p. 171–179, 2002.
- COOK, R. F.; LEROUX, C.; ISSEL, C. J. Equine infectious anemia and equine infectious anemia virus in 2013: A review. **Veterinary Microbiology**, v. 167, n. 1–2, p. 181–204, 2013.
- CURSINO, A. E.; VILELA, A. P. P.; FRANCO-LUIZ, A. P. M.; DE OLIVEIRA, J. G.; NOGUEIRA, M. F.; JÚNIOR, J. P. A.; DE AGUIAR, D. M.; KROON, E. G. Equine infectious anemia virus in naturally infected horses from the Brazilian Pantanal. **Archives of Virology**, v. 163, n. 9, p. 2385–2394, 2018.
- DONG, J.-B.; ZHU, W.; COOK, F. R.; GOTO, Y.; HORII, Y.; HAGA, T. Development of a nested PCR assay to detect equine infectious anemia proviral DNA from peripheral blood of naturally infected horses. **Archives of Virology**, v. 157, n. 11, p. 2105–2111, 2012.
- GARTON, N. J.; GILLERON, M.; BRANDO, T.; DAN, H. H.; GIGUÈRE, S.; PUZO, G.; PRESCOTT, J. F.; SUTCLIFFE, I. C. A novel lipoarabinomannan from the equine pathogen *Rhodococcus equi*: Structure and effect on macrophage cytokine production. **Journal of Biological Chemistry**, v. 277, n. 35, p. 31722–31733, 2002.

HINES, R.; MAURY, W. DH82 cells: a macrophage cell line for the replication and study of equine infectious anemia virus. **Journal of virological methods**, v. 95, n. 1–2, p. 47–56, 2001.

ISSEL, C. J.; COOK, R. F.; MEALEY, R. H.; HOROHOV, D. W. Equine infectious anemia in 2014: Live with it or eradicate it? **Veterinary Clinics of North America - Equine Practice**, v. 30, n. 3, p. 561–577, 2014.

ISSEL, C. J.; SCICLUNA, M. T.; COOK, S. J.; COOK, R. F.; CAPRIOLI, A.; RICCI, I.; ROSONE, F.; CRAIGO, J. K.; MONTELARO, R. C.; AUTORINO, G. L. Challenges and proposed solutions for more accurate serological diagnosis of equine infectious anaemia. **Veterinary Record**, v. 172, n. 8, p. 210, 2013.

KUHAR, U.; ZAVRŠNIK, J.; TOPLAK, I.; MALOVRH, T. Detection and molecular characterisation of equine infectious anaemia virus from field outbreaks in Slovenia. **Equine Veterinary Journal**, v. 46, n. 3, p. 386–391, 2014.

KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. **Molecular biology and evolution**, v. 33, n. 7, p. 1870–1874, 2016.

LANGEMEIER, J. L.; COOK, S. J.; COOK, R. F.; RUSHLOW, K. E.; MONTELARO, R. C.; ISSEL, C. J. Detection of equine infectious anemia viral RNA in plasma samples from recently infected and long-term inapparent carrier animals by PCR. **Journal of Clinical Microbiology**, v. 34, n. 6, p. 1481–1487, 1996.

LEROUX, C.; CADORÉ, J.-L.; MONTELARO, R. C. Equine Infectious Anemia Virus (EIAV): what has HIV's country cousin got to tell us? **Vet Res**, v. 35, p. 485–512, 2004.

NAGARAJAN, M. M.; SIMARD, C. Detection of horses infected naturally with equine infectious anemia virus by nested polymerase chain reaction. **Journal of Virological Methods**, v. 94, n. 1–2, p. 97–109, 2001.

NOGUEIRA, M. F.; OLIVEIRA, J. M.; SANTOS, C. J. S.; PETZOLD, H. V.; AGUIAR, D. M.; JULIANO, R. S.; REIS, J. K. P.; ABREU, U. G. P. Equine infectious anaemia in equids of Southern Pantanal, Brazil: Seroprevalence and evaluation of the adoption of a control programme. **Pesquisa Veterinaria Brasileira**, v. 37, n. 3, p. 227–233, 2017.

QUINLIVAN, M.; COOK, R. F.; CULLINANE, A. Real-time quantitative RT-PCR and PCR assays for a novel European field isolate of equine infectious anaemia virus based on sequence determination of the gag gene. **Veterinary Record**, v. 160, n. 18, p. 611–618, 2007.

REIS, J. K. P.; DINIZ, R. S.; HADDAD, J. P. A.; FERRAZ, I. B. F.; CARVALHO, A. F.; KROON, E. G.; FERREIRA, P. C. P.; LEITE, R. C. Recombinant envelope protein (rgp90) ELISA for equine infectious anemia virus provides comparable results to the agar gel immunodiffusion. **Journal of Virological Methods**, v. 180, n. 1–2, p. 62–67, 2012.

RICOTTI, S.; GARCIA, M. I.; VEAUTE, C.; BAILAT, A.; LUCCA, E.; COOK, R. F.; COOK, S. J.; SOUTULLO, A. Serologically silent, occult equine infectious anemia virus (EIAV) infections in horses. **Veterinary Microbiology**, v. 187, p. 41–49, 2016.

ROZEN, S.; SKALETISKY, H. Primer3 on the WWW for general users and for biologist programmers. **Methods Mol Biol**, v. 132, v. 1999/11/05, p. 365–386, 2000.

SCICLUNA, M. T.; AUTORINO, G. L.; NOGAROL, C.; RICCI, I.; FRONTOSO, R.; ROSONE, F.; NARDINI, R. Validation of an indirect ELISA employing a chimeric recombinant gag and env peptide for the serological diagnosis of equine infectious anemia. **Journal of Virological Methods**, v. 251, p. 111–117, 2018.

SCICLUNA, M. T.; ISSEL, C. J.; COOK, F. R.; MANNA, G.; CERSINI, A.; ROSONE, F.; FRONTOSO, R.; CAPRIOLI, A.; ANTONETTI, V.; AUTORINO, G. L. Is a diagnostic system based exclusively on agar gel immunodiffusion adequate for controlling the spread of equine infectious anaemia? **Veterinary Microbiology**, v. 165, n. 1–2, p. 123–134, 2013.

SUZUKI T., UEDA S. & SAMEJINA T. Enzyme-linked immunosorbent assay for diagnosis of equine infectious anaemia. **Vet. Microbiol.**, v. 7, p. 307–316, 1982.

6 Conclusões finais

Neste estudo foram obtidas duas sequências genômicas completas do VAIE em animais naturalmente infectados no Pantanal brasileiro, área endêmica para a doença. A estratégia da NGS desenvolvida foi capaz de obter leituras suficientes para uma sequência completa, sem pré-amplificação, da amostra de equídeos de RNA. Com base em alinhamentos de sequência e análise da árvore filogenética, BRA1 e BRA2 não estão intimamente relacionados com qualquer outra linhagem relatada até o momento, formando um grupo monofilético separado. A análise molecular dos genes mostrou alta variação de nucleotídeos e aminoácidos em LTR e *env*, e genes já relatados como conservados *pol* e *gag* com menor variação. Com a análise do gene do envelope, quasispécies virais diferentes puderam ser observados em ambos equinos.

Baseado nos dois genomas obtidos, um qPCR amplificando o gene *tat* de VAIE foi padronizado. A análise de 255 equinos com os testes diagnósticos disponíveis (IDGA, ELISA, Semi Nested PCR e o qPCR descrito neste trabalho) mostrou que nenhum teste possui capacidade de detectar todos os animais positivos. Um número significativo de equinos infectados pela VAIE pode escapar da detecção se apenas o teste AGID for realizado (padrão ouro). Para realmente implementar um plano de controle efetivo e minimizar a chance de manter animais positivos em meio a uma fazenda controlada, sugerimos a implementação dos testes disponíveis em uma sequência de diagnóstico.

7 Referências bibliográficas

- AMODEO, P.; MORELLI, M. A. C.; OSTUNI, A.; BATTISTUZZI, G.; BAVOSO, A. Structural features in EIAV NCp11: A lentivirus nucleocapsid protein with a short linker. **Biochemistry**, v. 45, n. 17, p. 5517–5526, 2006.
- ANDRAKE, M. D.; SKALKA, A. M. Retroviral Integrase: Then and Now. **Annual Review of Virology**, v. 2, n. 1, p. 241–264, 2015.
- BACCAM, P.; THOMPSON, R. J.; LI, Y.; SPARKS, W. O.; BELSHAN, M.; DORMAN, K. S.; WANNEMUEHLER, Y.; OAKS, J. L.; CORNETTE, J. L.; CARPENTER, S. Subpopulations of Equine Infectious Anemia Virus Rev Coexist In Vivo and Differ in Phenotype. **Journal of virology**, v. 77, n. 22, p. 12122–12131, 2003.
- BALL, J. M.; RUSHLOW, K. E.; ISSEL, C. J.; MONTELARO, R. C. Detailed mapping of the antigenicity of the surface unit glycoprotein of equine infectious anemia virus by using synthetic peptide strategies. **Journal of virology**, v. 66, n. 2, p. 732–742, 1992.
- BARROS, A. T. M. Seasonality and relative abundance of Tabanidae (Diptera) captured on horses in the Pantanal, Brazil. **Mem. Inst. Oswaldo Cruz**, v. 96, n. 7, p. 917-923, 2001.
- BARROS, T.; FOIL, L. Seasonal occurrence and relative abundance of Tabanidae (Diptera) from the Pantanal region. **Mem Entomol. Internl.**, v.14, p. 387-396, 1999.
- BELSHAN, M.; BACCAM, P.; OAKS, J. L.; SPONSELLER, B. A.; MURPHY, S. C.; CORNETTE, J.; CARPENTER, S. Genetic and biological variation in equine infectious anemia virus Rev correlates with variable stages of clinical disease in an experimentally infected pony. **Virology**, v. 279, n. 1, p. 185–200, 2001.
- BORGES, A. M. C. M.; SILVA, L. G.; NOGUEIRA, M. F.; OLIVEIRA, A. C. S.; SEGRI, N. J.; FERREIRA, F.; WITTER, R.; AGUIAR, D. M. Prevalence and risk factors for Equine Infectious Anemia in Poconé municipality, northern Brazilian Pantanal. **Research in Veterinary Science**, v. 95, n. 1, p. 76–81, 2013.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Instrução Normativa n. 45, de 15 de junho de 2004. Aprovar as normas para a prevenção e o controle da Anemia Infecciosa Equina - A.I.E. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 7 jul. 2004. Seção 1, p.7. Disponível em: <http://www.idaron.ro.gov.br/portal/legislacao/arquivos/exibir.ashx?arquivo=145&especie=Instrucao_Normativa&Num=45&ano=2004>. Acesso em: 23 nov. 2017.
- BRINDLEY, M. A.; MAURY, W. Endocytosis and a low-pH step are required for productive entry of equine infectious anemia virus. **Journal of virology**, v. 79, n. 23, p. 14482–14488, 2005.
- BRINDLEY, M. A.; ZHANG, B.; MONTELARO, R. C.; MAURY, W. An equine infectious

anemia virus variant superinfects cells through novel receptor interactions. **Journal of virology**, v. 82, n. 19, p. 9425–9432, 2008.

BOGERD, H. P. et al. Equine infectious anemia virus resists the antiretroviral activity of equine APOBEC3 proteins through a packaging-independent mechanism. **J. Virol.**, v.82, n. 23, p.11889–901, 2008.

BOLFA, P. F. et al. Oxidant-antioxidant imbalance in horses infected with equine infectious anaemia virus. **Vet. J.**, v.192, n. 3, p.449–54, 2012.

BUZON, M. J. et al. Deep Molecular Characterization of HIV-1 Dynamics under Suppressive HAART. **PLOS Pathogens**, v. 7, n. 10, e1002314, 2011.

CAPOMACCIO, S.; CAPPELLI, K.; COOK, R. F.; NARDI, F.; GIFFORD, R.; MARENZONI, M. L.; PASSAMONTI, F. Geographic structuring of global EIAV isolates: A single origin for New World strains? **Virus Research**, v. 163, n. 2, p. 656–659, 2012. a.

CAPOMACCIO, S.; WILLAND, Z. A.; COOK, S. J.; ISSEL, C. J.; SANTOS, E. M.; REIS, J. K. P.; COOK, R. F. Detection, molecular characterization and phylogenetic analysis of full-length equine infectious anemia (EIAV) gag genes isolated from Shackleford Banks wild horses. **Veterinary Microbiology**, v. 157, n. 3–4, p. 320–332, 2012. b.

CAPPELLI, K.; CAPOMACCIO, S.; COOK, F. R.; FELICETTI, M.; MARENZONI, M. L.; COPPOLA, G.; VERINI-SUPPLIZI, A.; COLETTI, M.; PASSAMONTI, F. Molecular detection, epidemiology, and genetic characterization of novel european field isolates of equine infectious anemia virus. **Journal of Clinical Microbiology**, v. 49, n. 1, p. 27–33, 2011.

CAPPELLI, K.; COOK, R. F.; STEFANETTI, V.; PASSAMONTI, F.; AUTORINO, G. L.; SCICLUNA, M. T.; COLETTI, M.; VERINI SUPPLIZI, A.; CAPOMACCIO, S. Deep sequencing and variant analysis of an Italian pathogenic field strain of equine infectious anaemia virus. **Transboundary and Emerging Diseases**, v. 64, n. 6, p. 2104–2112, 2017.

CARROLL, R.; DERSE, D. Translation of equine infectious anemia virus bicistronic tat-rev mRNA requires leaky ribosome scanning of the tat CTG initiation codon. **Journal of virology**, v. 67, n. 3, p. 1433–40, 1993.

CARROLL, R.; MARTARANO, L.; DERSE, D. Identification of lentivirus tat functional domains through generation of equine infectious anemia virus/human immunodeficiency virus type 1 tat gene chimeras. **Journal of virology**, v. 65, n. 7, p. 3460–7, 1991.

CARVALHO, M.; KIRKLAND, M.; DERSE, D. Protein Interactions with DNA Elements in Variant Equine Infectious Anemia Virus Enhancers and Their Impact on Transcriptional Activity. **Journal of virology**, v. 67, n. 11, p. 6586–6595, 1993.

CHARNEAU, P.; ALIZON, M.; CLAVEL, F. A Second Origin of DNA Plus-Strand Synthesis Is Required for Optimal Human Immunodeficiency Virus Replication. **Journal of Virology**, v. 66, n. 5, p. 2814–2820, 1992.

CHUNG, C.; MEALEY, R. H.; MCGUIRE, T. C. Evaluation of high functional avidity CTL to Gag epitope clusters in EIAV carrier horses. **Virology**, v. 342, n. 2, p. 228–239, 2005.

COGGINS, L.; NORCROSS, N. L.; NUSBAUM, S. R. Diagnosis of equine infectious anemia by immunodiffusion test. **American Journal of Veterinary Research**, v. 33, n. 1, p. 11–18, 1972.

COOK, R. F.; COOK, S. J.; LI, F.; MONTELARO, R. C.; ISSEL, C. J. Development of a multiplex real-time reverse transcriptase-polymerase chain reaction for equine infectious anemia virus (EIAV). **Journal of Virological Methods**, v. 105, n. 1, p. 171–179, 2002.

COOK, R. F.; LEROUX, C.; ISSEL, C. J. Equine infectious anemia and equine infectious anemia virus in 2013: A review. **Veterinary Microbiology**, v. 167, n. 1–2, p. 181–204, 2013.

COVALEDA, L.; GNO, B. T.; FULLER, F. J.; PAYNE, S. L. Identification of cellular proteins interacting with equine infectious anemia virus S2 protein. **Virus Research**, v. 151, n. 2, p. 235–239, 2010.

CRAIGO, J. K.; MONTELARO, R. C. EIAV envelope diversity: shaping viral persistence and encumbering vaccine efficacy. **Curr HIV Res**, v. 8, n. 1873–4251 (Electronic), p. 81–86, 2010.

CRAIGO, J. K.; MONTELARO, R. C. Lessons in AIDS vaccine development learned from studies of equine infectious, anemia virus infection and immunity. **Viruses**, v. 5, n. 12, p. 2963–2976, 2013.

CURSINO, A. E.; VILELA, A. P. P.; FRANCO-LUIZ, A. P. M.; DE OLIVEIRA, J. G.; NOGUEIRA, M. F.; JÚNIOR, J. P. A.; DE AGUIAR, D. M.; KROON, E. G. Equine infectious anemia virus in naturally infected horses from the Brazilian Pantanal. **Archives of Virology**, v. 163, n. 9, p. 2385–2394, 2018.

DAUTER, Z.; PERSSON, R.; ROSENGREN, A. M.; NYMAN, P. O.; WILSON, K. S.; CEDERGREN-ZEPPEZAUER, E. S. Crystal structure of dUTPase from equine infectious anaemia virus; Active site metal binding in a substrate analogue complex. **Journal of Molecular Biology**, v. 285, n. 2, p. 655–673, 1999.

DAVIES, J. F.; HOSTOMSKA, Z.; HOSTOMSKY, Z.; JORDAN, S. R.; MATTHEWS, D. A. Crystal Structure of the Ribonuclease H Domain of HIV-1 Reverse Transcriptase. **Science**, v. 252, p. 250–253, 1991.

DONG, J. B.; ZHU, W.; COOK, F. R.; GOTO, Y.; HORII, Y.; HAGA, T. Development of a nested PCR assay to detect equine infectious anemia proviral DNA from peripheral blood of naturally infected horses. **Archives of Virology**, v. 157, n. 11, p. 2105–2111, 2012.

DONG, J. B.; ZHU, W.; COOK, F. R.; GOTO, Y.; HORII, Y.; HAGA, T. Identification of a novel equine infectious anemia virus field strain isolated from feral horses in southern Japan. **Journal of General Virology**, v. 94, p. 360–365, 2013.

DONG, J.; COOK, F. R.; ZHU, W. Equine infectious anemia virus in Japan: Viral isolates V70 and V26 are of North American not Japanese origin. **Veterinary Microbiology**, v. 174, n. 1–2, p. 276–278, 2014.

DUNN, B. M.; GOODENOW, M. M.; GUSTCHINA, A.; WLODAWER, A. Retroviral proteases. **Genome Biology**, v. 3, n. 4, p. 1–7, 2002.

ELLERMAN, C.; BANG, O. Experimentelle Leukämie bei Hühnern. **Zentralbl. Bakteriол. Parasitenkd. Infectionskr. Hyg. Abt. Orig.** v. 46, p. 595–609, 1908.

FAGERNESS, A. J.; FLAHERTY, M. T.; PERRY, S. T.; JIA, B.; PAYNE, S. L.; FULLER, F. J. The S2 accessory gene of equine infectious anemia virus is essential for expression of disease in ponies. **Virology**, v. 349, n. 1, p. 22–30, 2006.

FIELDS, B. N.; KNIPE, D. M.; HOWLEY, P. M. **Fields virology**. Philadelphia: Ed. Wolters Kluwer Health/Lippincott Williams & Wilkins, 5th ed, 2007.

FLORES, E. F. **Virologia Veterinária**. Santa Maria: Ed. Da UFSM. 2007. 888p.

FOIL L. D., et al. Bloodmeal residues on mouthparts of *Tabanus fuscicostatus* (Diptera: Tabanidae) and the potential for mechanical transmission of pathogens. **J. Med. Entomol.**, v.24, n. 6, p. 6–613, 1987.

FOIL, L. D. et al. Mechanical transmission of equine infectious anemia by deer flies. (*Chrisops flavidus*) and stable flies (*Stomoxys calcitrans*). **Am. J. Vet. Res.**, v. 44, n. 1, p. 155–156, 1983.

GANSER-PORNILLOS, B. K.; YEAGER, M.; SUNDQUIST, W. I. The structural biology of HIV assembly. **Current Opinion in Structural Biology**, v. 18, n. 2, p. 203–217, 2008.

GAO, X.; JIANG, C. G.; WANG, X. F.; LIN, Y. Z.; MA, J.; HAN, X. E.; ZHAO, L. P.; SHEN, R. X.; XIANG, W. H.; ZHOU, J. H. Reverse mutation of the virulence-associated S2 gene does not cause an attenuated equine infectious anemia virus strain to revert to pathogenicity. **Virology**, v. 443, n. 2, p. 321–328, 2013.

GENDELMAN, H. E.; NARAYAN, O.; KENNEDY-STOSKOPF, S.; KENNEDY, P.; GHOTBI, Z.; CLEMENTS, J.; STANLEY, J.; PEZESHKPOUR, G. Tropism of sheep lentiviruses for monocytes: susceptibility to infection and virus gene expression increase during maturation of monocytes to macrophages. **Journal of virology**, v. 58, p. 67–74, 1986.

GRUND, C. H.; LECHMAN, E. R.; ISSEL, C. J.; MONTELARO, R. C.; RUSHLOW, K. E. Lentivirus cross-reactive determinants present in the capsid protein of equine infectious anaemia virus. **Journal of General Virology**, v. 75, n. 3, p. 657–662, 1994.

GUERREIRO, M. G. et al. Simpósio sobre anemia infecciosa equina. **Boletim do Instituto de Pesquisa Veterinária "Desidério Finamor"**, Porto Alegre, n. 1/2, p. 3–4, 1968.

GUSTCHINA, A.; KERVINEN, J.; POWELL, D. J.; ZDANOV, A.; KAY, J.; WLODAWER, A. Structure of equine infectious anemia virus proteinase complexed with an inhibitor. **Protein Science**, v. 5, n. 8, p. 1453–1465, 1996.

HAN, G.Z., WOROBEY, M., 2012. An endogenous foamy-like viral element in the coelacanth genome. **PLoS Pathogens**, v. 8, n. 6, e1002790.

HARRIS, M. E.; GONTAREK, R. R.; DERSE, D.; HOPE, T. J. Differential requirements for alternative splicing and nuclear export functions of equine infectious anemia virus Rev protein. **Mol Cell Biol**, v. 18, n. 7, p. 3889–3899, 1998.

HATANAKA, H.; IOURIN, O.; RAO, Z.; FRY, E.; KINGSMAN, A.; STUART, D. I. Structure of Equine Infectious Anemia Virus Matrix Protein. **Journal of Virology**, v. 76, n. 4, p. 1876–1883, 2002.

IBGE 2016. Banco de Dados Agregados. Tabela 3939: efetivo dos rebanhos, por tipo de rebanho. Sistema Instituto Brasileiro de Geografia e Estatística (IBGE) de Recuperação Automática - SIDRA. Disponível em <<https://sidra.ibge.gov.br/tabela/3939#resultado>>. Acesso em 10 out. 2017.

ICTV 2018. International Committee on Taxonomy of Viruses. Disponível em <<https://talk.ictvonline.org/taxonomy/>>. Acesso em 14 dez. 2018.

ISSEL, C. J.; COOK, R. F. A Review of Techniques for the Serologic Diagnosis of Equine Infectious Anemia. **Journal of Veterinary Diagnostic Investigation**, v. 5, n. 1, p. 137–141, 1993.

ISSEL, C. J.; COOK, R. F.; MEALEY, R. H.; HOROHOV, D. W. Equine infectious anemia in 2014: Live with it or eradicate it? **Veterinary Clinics of North America - Equine Practice**, v. 30, n. 3, p. 561–577, 2014.

ISSEL, C. J.; FOIL, L. D. Studies on equine infectious anemia virus transmission by insects. **J. Am. Vet. Med. Assoc.**, v. 184, n. 3, p. 7-293, 1984.

ISSEL, C. J.; SCICLUNA, M. T.; COOK, S. J.; COOK, R. F.; CAPRIOLI, A.; RICCI, I.; ROSONE, F.; CRAIGO, J. K.; MONTELARO, R. C.; AUTORINO, G. L. Challenges and proposed solutions for more accurate serological diagnosis of equine infectious anaemia. **Veterinary Record**, v. 172, n. 8, p. 210, 2013.

JIN, S.; ZHANG, B.; WEISZ, O. A.; MONTELARO, R. C. Receptor-Mediated Entry by Equine Infectious Anemia Virus Utilizes a pH-Dependent Endocytic Pathway. **Journal of Virology**, v. 79, n. 23, p. 14489–14497, 2005.

KATZOURAKIS, A.; TRISTEM, M.; PYBUS, O. G.; GIFFORD, R. J. Discovery and analysis of the first endogenous lentivirus. **Proceedings of the National Academy of Sciences**, v. 104, n. 15, p. 6261–6265, 2007.

KOBAYASHI, K.; KONO, Y. Propagation and titration of equine infectious anemia virus in

horse leukocyte culture. **Natl. Inst. Anim. Health Q Tokyo**, v. 7, p. 8–20, 1967.

LEE, J.-H.; MURPHY, S. C.; BELSHAN, M.; SPARKS, W. O.; WANNEMUEHLER, Y.; LIU, S.; HOPE, T. J.; DOBBS, D.; CARPENTER, S. Characterization of functional domains of equine infectious anemia virus Rev suggests a bipartite RNA-binding domain. **Journal of Virology**, v. 80, n. 8, p. 3844–3852, 2006.

LEE, J. H.; CULVER, G.; CARPENTER, S.; DOBBS, D. Analysis of the EIAV Rev-Responsive Element (RRE) reveals a conserved RNA motif required for high affinity Rev binding in both HIV-1 and EIAV. **PLoS ONE**, v. 3, n. 6, 2008.

LEROUX, C.; CADORÉ, J.-L.; MONTELARO, R. C. Equine Infectious Anemia Virus (EIAV): what has HIV's country cousin got to tell us? **Vet Res**, v. 35, p. 485–512, 2004.

LEROUX, C.; CRAIGO, J. K.; ISSEL, C. J.; MONTELARO, R. C. Equine Infectious Anemia Virus Genomic Evolution in Progressor and Nonprogressor Ponies. **Journal of Virology**, v. 75, n. 10, p. 4570–4583, 2001.

LEROUX, C.; ISSEL, C. J.; MONTELARO, R. C. Novel and dynamic evolution of equine infectious anemia virus genomic quasiespecies associated with sequential disease cycles in an experimentally infected pony. **Journal of virology**, v. 71, n. 12, p. 9627–39, 1997.

LI, F.; CRAIGO, J. K.; HOWE, L.; STECKBECK, J. D.; COOK, S.; ISSEL, C.; MONTELARO, R. C. A Live Attenuated Equine Infectious Anemia Virus Proviral Vaccine with a Modified. **Journal of Virology**, v. 77, n. 13, p. 7244–7253, 2003.

LI, F.; LEROUX, C.; CRAIGO, J. K.; COOK, S. J.; ISSEL, C. J.; MONTELARO, R. C. The S2 gene of equine infectious anemia virus is a highly conserved determinant of viral replication and virulence properties in experimentally infected ponies. **Journal of Virology**, v. 74, n. 1, p. 573–9, 2000.

LI, F.; PUFFER, B. A.; MONTELARO, R. C. The S2 gene of equine infectious anemia virus is dispensable for viral replication in vitro. **J. Virol.**, v. 72, n. 10, p. 8344–8348, 1998.

LICHTENSTEIN, D. L.; CRAIGO, J. K.; LEROUX, C.; RUSHLOW, K. E.; COOK, R. F.; COOK, S. J.; ISSEL, C. J.; MONTELARO, R. C. Effects of long terminal repeat sequence variation on equine infectious anemia virus replication in vitro and in vivo. **Virology**, v. 263, n. 2, p. 408–417, 1999.

LIGNÉ, M., Mémoire et observations sur une maladie de sang, connue sous le nom d'anémie hydroémie, cachexie acquise du cheval, **Rec. Med. Vet. Ec. Alfort**, v. 20, p. 30–44, 1843.

LIU, H. et al. Complete Sequence of Proviral DNA of Equine Infectious Anemia Virus Strain L(Liaoning). **Chinese Journal of Preventive Veterinary Medicine**. S1, 2000.

LIU, Q.; WANG, X.-F.; MA, J.; HE, X.-J.; WANG, X.-J.; ZHOU, J.-H. Characterization of Equine Infectious Anemia Virus Integration in the Horse Genome. **Viruses**, v. 7, n. 6, p.

3241–3260, 2015.

MATHEKA, H. D. et al. Purification and characterization of equine infectious anemia virus. **Archives of Virology**, v. 51, p. 107–114, 1976.

MATSRAFI, L.; AFLALO, E.; BLAIR, D. G.; PRIEL, E. Isolation of an 11-kDa protein associated with the topoisomerase I activity from equine infectious anemia virus. **Biochemical and Biophysical Research Communications**, v. 220, n. 3, p. 1028–1035, 1996.

MAURY, W. Monocyte maturation controls expression of equine infectious anemia virus. **Journal of virology**, v. 68, n. 10, p. 6270–6279, 1994.

MAURY, W. Regulation of equine infectious anemia virus expression. **Journal of Biomedical Science**, v. 5, n. 1, p. 11–23, 1998.

MAURY, W.; PERRYMAN, S.; OAKS, J. L.; SEID, B. K.; CRAWFORD, T.; MCGUIRE, T.; CARPENTER, S. Localized sequence heterogeneity in the long terminal repeats of in vivo isolates of equine infectious anemia virus. **Journal of virology**, v. 71, n. 7, p. 4929–4937, 1997.

MAURY, W.; THOMPSON, R. J.; JONES, Q.; BRADLEY, S.; DENKE, T.; BACCAM, P.; SMAZIK, M.; OAKS, J. L. Evolution of the equine infectious anemia virus long terminal repeat during the alteration of cell tropism. **Journal of virology**, v. 79, n. 9, p. 5653–5664, 2005.

MOLIKI, J. L.; ROHWER, F.; DUTILH, B. E. Metagenomics and future perspectives in virus discovery. **Curr Opin Virol**, v. 2: p. 63-77, 2012.

NAGARAJAN, M. M.; SIMARD, C. Detection of horses infected naturally with equine infectious anemia virus by nested polymerase chain reaction. **Journal of Virological Methods**, v. 94, n. 1–2, p. 97–109, 2001.

NOGUEIRA, M. F.; OLIVEIRA, J. M.; SANTOS, C. J. S.; PETZOLD, H. V.; AGUIAR, D. M.; JULIANO, R. S.; REIS, J. K. P.; ABREU, U. G. P. Equine infectious anaemia in equids of Southern Pantanal, Brazil: Seroprevalence and evaluation of the adoption of a control programme. **Pesquisa Veterinaria Brasileira**, v. 37, n. 3, p. 227–233, 2017.

OIE. World Organization for Animal Health. Equine Infectious Anaemia. In: Manual of diagnostic tests and vaccines for terrestrial animals. Paris: OIE, p.1-6, 2013. Disponível em: <http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.05.06_EIA.pdf>. Acesso em 28 out. 2017.

OIE. World Organization for Animal Health. Equine Infectious Anaemia. World Animal Health Information Database (WAHIS) Interface. OIE, 2018. Disponível em: <http://www.oie.int/wahis_2/public/wahid.php/Wahidhome/Home>. Acesso em 28 nov. 2018.

PAYNE, S. L.; FULLER, F. J. Virulence determinants of equine infectious anemia virus. **Current HIV research**, v. 8, n. 1, p. 66–72, 2010.

PEARSON, J. E.; BECVAR, C. S.; MOTT, L. O. Evaluation of the immunodiffusions test for the diagnosis of equine infectious anemia. **Proc US Animal Health Association**, v. 74, p. 259-267, 1971.

PERRYMAN, L. E. et al. Decreased C3 levels in a chronic virus infection: equine infectious anemia. **J. Immunol.**, v. 106, n. 4, p. 1074–8, 1971.

PETROPOULOS, C. J. Retroviral taxonomy, protein structures, sequences, and genetic maps. In: COFFIN, J. M. (Ed.). **Retroviruses**. New York: Cold Spring Harbor Laboratory Press, 1997.

PUFFER, B. A.; PARENT, L. J.; WILLS, J. W.; MONTELARO, R. C. Equine infectious anemia virus utilizes a YXXL motif within the late assembly domain of the Gag p9 protein. **Journal of virology**, v. 71, n. 9, p. 6541–6, 1997.

PUFFER, B. A.; WATKINS, S. C.; MONTELARO, R. C. Equine Infectious Anemia Virus Gag Polyprotein Late Domain Specifically Recruits Cellular AP-2 Adapter Protein Complexes during Virion Assembly. v. 72, n. 12, p. 10218–10221, 1998.

QUINLIVAN, M.; COOK, F.; KENNA, R.; CALLINAN, J. J.; CULLINANE, A. Genetic characterization by composite sequence analysis of a new pathogenic field strain of equine infectious anemia virus from the 2006 outbreak in Ireland. **Journal of General Virology**, v. 94, p. 612–622, 2013.

QUINLIVAN, M.; COOK, R. F.; CULLINANE, A. Real-time quantitative RT-PCR and PCR assays for a novel European field isolate of equine infectious anaemia virus based on sequence determination of the gag gene. **Veterinary Record**, [s. l.], v. 160, n. 18, p. 611–618, 2007.

REIS, J. K. P.; CRAIGO, J. K.; COOK, S. J.; ISSEL, C. J.; MONTELARO, R. C. Characterization of EIAV LTR variability and compartmentalization in various reservoir tissues of long-term inapparent carrier ponies. **Virology**, v. 311, n. 1, p. 169–180, 2003.

REIS, J. K. P.; DINIZ, R. S.; HADDAD, J. P. A.; FERRAZ, I. B. F.; CARVALHO, A. F.; KROON, E. G.; FERREIRA, P. C. P.; LEITE, R. C. Recombinant envelope protein (rgp90) ELISA for equine infectious anemia virus provides comparable results to the agar gel immunodiffusion. **Journal of Virological Methods**, v. 180, n. 1–2, p. 62–67, 2012.

REIS, J. K. P.; LEITE, R. C. Anemia Infecciosa Equina. In: MEGID, J.; RIBEIRO, M. G.; PAES, A. C. **Doenças Infecciosas em Animais de Produção e de Companhia**. Rio de Janeiro: Roca, 2016; p. 545-553.

RICE, N. R.; HENDERSON, L. E.; SOWDER, R. C.; COPELAND, T. D.; OROSZLAN, S.; EDWARDS, J. F. Synthesis and processing of the transmembrane envelope protein of equine infectious anemia virus. **Journal of Virology**, v. 64, n. 8, p. 3770–3778, 1990.

ROSSMANITH W, HORVATH E. A western blot test for the serological diagnosis of equine

infectious anemia. **Zentralbl Veterinarmed B**, v. 36, n. 1, p. 49-56, 1989.

ROSSI, M. M.; TERNES, S. Estimativa do Valor de R₀ pelo Tamanho Final da Doença, a : Estudo de caso da Anemia Infecciosa Equina no Pantanal. **X Congresso Brasileiro de Agroinformática**, 2015.

ROSSI, M.; TERNES, S.; NOGUEIRA, M. F. 2015. Considerações sobre o R₀ da Anemia Infecciosa Equina a partir dos agentes infecciosos inseto-vetor “Mutuca” e “seringas contaminadas”. **Comunicado técnico 121**. ISSN 1677-8464. Embrapa - Campinas, SP, 2015.

ROUS, P. Transmissible avian neoplasm (sarcoma of the common fowl). **Journal of Experimental Medicine**, v. 12, p. 696–705, 1910.

RUNGROTMONGKOL, T.; MULHOLLAND, A. J.; HANNONGBUA, S. Active site dynamics and combined quantum mechanics/molecular mechanics (QM/MM) modelling of a HIV-1 reverse transcriptase/DNA/dTTP complex. **Journal of Molecular Graphics and Modelling**, v. 26, n. 1, p. 1–13, 2007.

SCICLUNA, M. T.; AUTORINO, G. L.; NOGAROL, C.; RICCI, I.; FRONTOSO, R.; ROSONE, F.; NARDINI, R. Validation of an indirect ELISA employing a chimeric recombinant gag and env peptide for the serological diagnosis of equine infectious anemia. **Journal of Virological Methods**, v. 251, p. 111–117, 2018.

SCICLUNA, M. T.; ISSEL, C. J.; COOK, F. R.; MANNA, G.; CERSINI, A.; ROSONE, F.; FRONTOSO, R.; CAPRIOLI, A.; ANTONETTI, V.; AUTORINO, G. L. Is a diagnostic system based exclusively on agar gel immunodiffusion adequate for controlling the spread of equine infectious anaemia? **Veterinary Microbiology**, v. 165, n. 1–2, p. 123–134, 2013.

SILVA, J. dos S. V. DA; ABDON, M. D. M. Delimitação do Pantanal Brasileiro e suas sub-regiões. **Pesq. Agropec. Bras.**, v. 33, p. 1703–1711, 1998.

SILVA, R. A. M. S.; ABREU, U. G. P. De; BARROS, A. T. M. De. Anemia Infecciosa Equina : Epizootiologia , Prevenção e Controle no Pantanal. **Embrapa: Pantanal - Circular Técnica**, 29, p. 1–32, 2001. Disponível em <https://www.infoteca.cnptia.embrapa.br/handle/doc/807376>. Acesso em 28/09/2018.

SILVA, R. A. M. S. et al. Equine viral diseases in the Pantanal, Brazil: studies carried out from 1990 to 1995. **Revue d'Élevage et de Médecine Vétérinaire des Pays Tropicaux**, v. 52, p. 9-12, 1999a.

SILVA, R. A. M. S. et al. Swamp Fever in wild horses from the Pantanal, Brazil. **Revue D'Élevage et Médecine Vétérinaire des Pays Tropicaux**. v.52, p.99-101, 1999b.

SILVA, R. A. M. S. et al. Programa de prevenção e controle da anemia infecciosa equina no Pantanal Sul-Mato-Grossense. **Documentos 68**, Embrapa Pantanal, Corumbá. 2004, 17p.

SIMON-LORIERE, E.; HOLMES, E. C. Why do RNA viruses recombine? **Nature Reviews Microbiology**, v. 9, n. 8, p. 617–626, 2011.

STAKE, M. S.; BANN, D. V.; KADDIS, R. J.; PARENT, L. J. Nuclear trafficking of retroviral RNAs and gag proteins during late steps of replication. **Viruses**, v. 5, n. 11, p. 2767–2795, 2013.

STEIN, C. D.; LOTZE, J. C.; MOTT, L. O. Transmission of equine infectious anemia by the stablefly, *Stomoxys calcitrans*, the horsefly, *Tabanus sulcifrons* (Macquart), and by injection of minute amounts of virus. **Am J Vet Res**, v. 3, p. 183–93, 1942.

SUÑÉ, C.; GOLDSTROHM, A. C.; PENG, J.; PRICE, D. H.; GARCIA-BLANCO, M. A.; OLIVEIRA, F. G.; COOK, R. F.; NAVES, J. H. F.; OLIVEIRA, C. H. S.; DINIZ, R. S.; FREITAS, F. J. C.; LIMA, J. M.; SAKAMOTO, S. M.; LEITE, R. C.; ISSEL, C. J.; REIS, J. K. P. An in vitro transcription system that recapitulates equine infectious anemia virus Tat-mediated inhibition of human immunodeficiency virus type 1 Tat activity demonstrates a role for positive transcription elongation factor b and associated proteins in th. **Virology**, v. 274, n. 2, p. 356–366, 2000.

SUZUKI, T.; UEDA, S.; SAMEJIMA, T. Enzyme-linked immunosorbent assay for diagnosis of equine infectious anaemia. **Vet. Microbiol.**, v. 7, p. 307–316, 1982.

TANG, H.; KUHEN, K. L.; WONG-STAAL, F. LENTIVIRUS REPLICATION AND REGULATION. **Annu. Rev. Genet.**, v. 33, p. 133–170, 1999.

TÖZSÉR, J. Comparative studies on retroviral proteases: Substrate specificity. **Viruses**, v. 2, n. 1, p. 147–165, 2010.

WANG, L.; TONG, G.; LIU, H.; YANG, Z.; QIU, H.; KONG, X.; WANG, M. Proviral genomic sequence analysis of Chinese donkey leukocyte attenuated equine infectious anemia virus vaccine and its parental virus strain Liaoning. **Science in China, Series C: Life Sciences**, v. 45, n. 1, 2002.

VALLEE, H.; CARRE, H. Sur la natur infectieuse de l'anémie du cheval. **CR Acad Sci**. v. 139, p. 331–3, 1904.

WANG, X. F.; LIN, Y. Z.; LI, Q.; LIU, Q.; ZHAO, W. W.; DU, C.; CHEN, J.; WANG, X.; ZHOU, J. H. Genetic Evolution during the development of an attenuated EIAV vaccine. **Retrovirology**, v. 13, n. 1, p. 1–13, 2016.

WEILAND, F et al. Electron microscopic studies on equine infectious anemia virus (EIAV). **Archives of virology**. v. 55, p. 335–340, 1977.

YU, H.; JETZT, A. E.; RON, Y.; PRESTON, B. D.; DOUGHERTY, J. P. The Nature of Human Immunodeficiency Virus Type 1 Strand Transfers. **Journal of Biological Chemistry**, v. 273, n. 43, p. 28384–28391, 1998.

ZHANG, B. S.; JIN, S.; JIN, J.; LI, F.; MONTELARO, R. C. A tumor necrosis factor receptor family protein serves as a cellular receptor for the macrophage-tropic equine lentivirus. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102,

n. 28, p. 9918–9923, 2005.

ZHANG, Z.; MA, J.; ZHANG, X.; SU, C.; YAO, Q.; WANG, X. Equine Infectious Anemia Virus Gag Assembly and Export Are Directed by Matrix Protein through trans -Golgi Networks and Cellular Vesicles. **Journal of Virology**, v. 90, n. 4, p. 1824–1838, 2016.

ZHENG Y.H. et al. In vivo dynamics of equine infectious anemia viruses emerging during febrile episodes: insertions/duplications at the principal neutralizing domain. **J Virol**. v. 71, n. 7, p. 5031–9, 1997.

8 Anexos

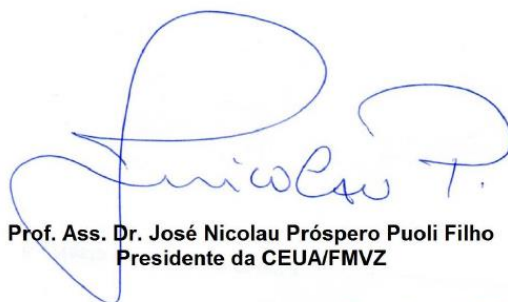
Anexo A: Atestado de Protocolo nº 195/2012-CEUA de aprovação do Projeto de Pesquisa.



A T E S T A D O

Atesto para os devidos fins, que o Projeto de Pesquisa **"Anemia infecciosa equina no Pantanal brasileiro: caracterização do agente, diagnóstico molecular, avaliação de práticas de manejo e modelagem quantitativa"**, Protocolo nº 195/2012-CEUA, do Professor **João Pessoa Araújo Júnior**, está de acordo com os Princípios Éticos na Experimentação Animal e foi aprovado pela Comissão de Ética no Uso de Animais (CEUA) desta Faculdade, e foi aprovado na reunião do dia 03 de outubro de 2012.

Botucatu, 15 de agosto de 2018.



Prof. Ass. Dr. José Nicolau Próspero Puoli Filho
Presidente da CEUA/FMVZ



UNESP – Faculdade de Medicina Veterinária e Zootecnia
Seção Técnica Acadêmica
Rua Prof.Dr. Walter Mauricio Corrêa, s/n
UNESP - Câmpus de Botucatu/SP - Cep 18618-681
☎ (14) 3880.2175 – ✉ sta@fmvz.unesp.br – 🌐 www.fmvz.unesp.br



Anexo B: Declaração para o Atestado de Protocolo nº 195/2012-CEUA, integrando os projetos realizados.



UNIVERSIDADE ESTADUAL PAULISTA
CAMPUS DE BOTUCATU
INSTITUTO DE BIOCÊNCIAS
Departamento de Microbiologia e Imunologia

Botucatu, 15 de agosto de 2018

DECLARAÇÃO

Declaro para os devidos fins que o projeto intitulado **"CARACTERIZAÇÃO MOLECULAR DO VÍRUS DA ANEMIA INFECCIOSA EQUINA NO PANTANAL E PADRONIZAÇÃO DE qPCR PARA DIAGNÓSTICO"** é parte integrante do projeto de pesquisa **"Anemia Infecciosa Equina no Pantanal brasileiro: caracterização do agente, avaliação de práticas de manejo e modelagem quantitativa"** que foi aprovado pelo comitê de ética no uso de animais da FMVZ - UNESP em 03 de outubro de 2012, Protocolo nº 195/2012-CEUA.

Atenciosamente,

Prof. Dr. João Pessoa Araújo Junior

Coordenador dos projetos.

Coordenador do Programa de PG em Biotecnologia IBB - UNESP

9 Apêndices

Appendix A: Table of SNPs from BRA1. Genome position number is given as BRA1 consensus reference.

Genome position	Change	Coverage	Variant Frequency	Variant P-Value	Amino Acid Change	Codon Change	Protein Effect	CDS
327	T -> A	201	25.90%	1.00E-123				LTR
502	A -> G	162	28.40%	8.40E-112	K -> R	AAG -> AGG	Substitution	Gag
1,067	G -> A	123	29.30%	3.80E-90		AAG -> AAA	None	Gag
1,799	G -> A	241	37.30%	2.50E-239		GAG -> GAA	None	Gag
1,830	A -> G	263	34.20%	1.10E-216	N -> D	AAC -> GAC	Substitution	Gag
2,024	T -> C	471	37.80%	0.00E+00		TTA -> CTA	None	Pol
2,046	A -> G	387	45.00%	0.00E+00	K -> R	AAG -> AGG	Substitution	Pol
2,155	G -> T	347	39.80%	0.00E+00	L -> F	TTG -> TTT	Substitution	Pol
2,157	A -> G	329	48.60%	0.00E+00	K -> R	AAG -> AGG	Substitution	Pol
2,308	T -> C	125	27.20%	3.40E-86		ATT -> ATC	None	Pol
2,364	G -> A	191	34.60%	6.80E-173	R -> K	AGA -> AAA	Substitution	Pol
2,462	A -> T	649	36.70%	0.00E+00	I -> L	ATA -> TTA	Substitution	Pol
2,478	C -> T	423	33.60%	0.00E+00	A -> V	GCT -> GTT	Substitution	Pol
2,539	A -> G	433	46.70%	0.00E+00		GGA -> GGG	None	Pol
2,952	A -> G	465	40.90%	0.00E+00	N -> S	AAT -> AGT	Substitution	Pol
3,144	A -> G	399	46.60%	0.00E+00	K -> R	AAA -> AGA	Substitution	Pol
3,184	G -> A	463	25.10%	1.90E-251		CAG -> CAA	None	Pol
3,482	A -> G	1,307	48.90%	0.00E+00	I -> V	ATA -> GTA	Substitution	Pol
3,699	G -> A	1,149	28.00%	0.00E+00	R -> K	AGG -> AAG	Substitution	Pol
3,727	C -> T	1,131	34.80%	0.00E+00		GAC -> GAT	None	Pol
3,985	C -> T	1,226	50.70%	0.00E+00		TAC -> TAT	None	Pol
4,099	A -> G	1,753	47.70%	0.00E+00		GGA -> GGG	None	Pol
4,249	G -> A	1,709	26.90%	0.00E+00		CAG -> CAA	None	Pol
4,273	T -> C	1,589	35.40%	0.00E+00		TCT -> TCC	None	Pol
4,381	A -> G	871	48.90%	0.00E+00		GCA -> GCG	None	Pol
4,693	A -> G	2,881	50.20%	0.00E+00		CAA -> CAG	None	Pol
5,881	G -> A	2,164	32.20%	0.00E+00	R -> K	AGA -> AAA	Substitution	Env
5,885	C -> G	2,104	32.70%	0.00E+00	N -> K	AAC -> AAG	Substitution	Env
5,912	A -> T	1,847	32.00%	0.00E+00	K -> N	AAA -> AAT	Substitution	Env
5,929	C -> A	2,105	97.10%	0.00E+00	T -> N	ACC -> AAC	Substitution	Env
5,941	A -> C	2,259	98.80%	0.00E+00	H -> P	CAT -> CCT	Substitution	Env
5,943	C -> T	2,283	99.20%	0.00E+00	H -> Y	CAT -> TAT	Substitution	Env
7,480	A -> G	16,497	54.30%	0.00E+00	N -> S	AAT -> AGT	Substitution	Env/Rev

Appendix B: Alignment of env clone sequences of BRA1 and BRA2. The BRA1 sequence was used as reference to the alignment. Inclusions and deletions are marked in yellow. Hypervariable region V3 is highlighted in green and V4 in pink.

	1	10	20	30	40	50	60
BRA1							
BRA1 Sequence 3	ATGGTCAGCATCGCGTTCTATGGGGGAATCCCAGGGGGAATCTCGACCCCTATTATTCAA						
BRA2 sequence 4.1	-----						
BRA1 Sequence 1	-----						
BRA2 sequence 5.1	-----						
BRA1 Sequence 2	-----						
BRA2 sequence 2.1	-----						
BRA2 sequence 2.2	-----						
BRA2 sequence 4.2	-----						
BRA2 sequence 5.2	-----						
BRA2 sequence 3.2	-----						
BRA2 sequence 6.2	-----						
BRA1 Sequence 1	-----						
BRA2 sequence 1.1	-----						
BRA1 Sequence 4	-----						
BRA2 sequence 3.1	-----						
BRA1 Sequence 2	-----						
BRA1	TCTTCAGAAGAAGCTAGTCACACCAAGGGGGATACAATATTTAAACCCCTATTGTTCTGAT						
BRA1 Sequence 3	TCTTCAGAAGAAGCTAGTCACACCAAGGGGGATACAATATTTAAACCCCTATTGTTCTGAT						
BRA2 sequence 4.1	-----						
BRA1 Sequence 1	-----						
BRA2 sequence 5.1	-----						
BRA1 Sequence 2	-----						
BRA2 sequence 2.1	-----						
BRA2 sequence 2.2	-----						
BRA2 sequence 4.2	-----						
BRA2 sequence 5.2	-----						
BRA2 sequence 3.2	-----						
BRA2 sequence 6.2	-----						
BRA1 Sequence 1	-----						
BRA2 sequence 1.1	-----						
BRA1 Sequence 4	-----						
BRA2 sequence 3.1	-----						
BRA1 Sequence 2	-----						
BRA1	GAAAAAAGTAACAGGAATAAAACAATGGCGGAAGAAAGAGATTCAAGAGAACGAAGAGAA						
BRA1 Sequence 3	GAAAAAAGTAACAGGAATAAAACAATGGCGGAAGAAAGAGATTCAAGAGAACGAAGAGAA						
BRA2 sequence 4.1	-----						
BRA1 Sequence 1	-----						
BRA2 sequence 5.1	-----						
BRA1 Sequence 2	-----						
BRA2 sequence 2.1	-----						
BRA2 sequence 2.2	-----						
BRA2 sequence 4.2	-----						
BRA2 sequence 5.2	-----						
BRA2 sequence 3.2	-----						
BRA2 sequence 6.2	-----						
BRA1 Sequence 1	-----						
BRA2 sequence 1.1	-----						
BRA1 Sequence 4	-----						
BRA2 sequence 3.1	-----						
BRA1 Sequence 2	-----						
BRA1	GCGTACCTGAAAGAAGATGCAGAAGAAAATGCAAAGAAGAGAAAACAATGACTGGTGGAAA						

BRA1	Sequence	3	CGGTACCTGAAAGAAGATGCAGAAGAAAATCCAAAGAAGAGAAACAATGACTGGTGGAAA
BRA2	sequence	4.1	-----
BRA1	Sequence	1	-----
BRA2	sequence	5.1	-----
BRA1	Sequence	2	-----
BRA2	sequence	2.1	-----
BRA2	sequence	2.2	-----
BRA2	sequence	4.2	-----
BRA2	sequence	5.2	-----
BRA2	sequence	3.2	-----
BRA2	sequence	6.2	-----
BRA1	Sequence	1	-----
BRA2	sequence	1.1	-----
BRA1	Sequence	4	-----
BRA2	sequence	3.1	-----
BRA1	Sequence	2	-----

BRA1			ATAGGTATGTTTATGCTATGTTTAATGGGAGCAACAGGAGGAATGCTCTGGTGGTATGAG
BRA1	Sequence	3	ATAGGTATGTTTATGCTATGTTTAATGGGAGCAACAGGAGGAATACTCTGGTGGTATGAG
BRA2	sequence	4.1	-----
BRA1	Sequence	1	-----
BRA2	sequence	5.1	-----
BRA1	Sequence	2	-----
BRA2	sequence	2.1	-----
BRA2	sequence	2.2	-----
BRA2	sequence	4.2	-----
BRA2	sequence	5.2	-----
BRA2	sequence	3.2	-----
BRA2	sequence	6.2	-----
BRA1	Sequence	1	-----
BRA2	sequence	1.1	-----
BRA1	Sequence	4	-----
BRA2	sequence	3.1	-----
BRA1	Sequence	2	-----

BRA1			GATGTGGCACACCCACAATATATAGGATTGATTACTGTTGGTGGGAAATTGGAAGGATCT
BRA1	Sequence	3	GATGTGGCACACCCACAATATATAGGATTGATTACTGTTGGTGGGAAATTGGAAGGATCT
BRA2	sequence	4.1	-----
BRA1	Sequence	1	-----
BRA2	sequence	5.1	-----
BRA1	Sequence	2	-----
BRA2	sequence	2.1	-----
BRA2	sequence	2.2	-----
BRA2	sequence	4.2	-----
BRA2	sequence	5.2	-----
BRA2	sequence	3.2	-----
BRA2	sequence	6.2	-----
BRA1	Sequence	1	-----
BRA2	sequence	1.1	-----
BRA1	Sequence	4	-----
BRA2	sequence	3.1	-----
BRA1	Sequence	2	-----

BRA1			GGAATGACTAGTGCTATAGAATGTTGGGGT-AAGTTTCCTGGCTGTAGACCGTTTGATAA
BRA1	Sequence	3	GGAATGACTAGTGCTATAGAATGTTGGGGTNAAGTTTCCTGGCTGTAGACCGTTTGATAA
BRA2	sequence	4.1	-----
BRA1	Sequence	1	-----
BRA2	sequence	5.1	-----
BRA1	Sequence	2	-----
BRA2	sequence	2.1	-----
BRA2	sequence	2.2	-----
BRA2	sequence	4.2	-----

```

BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 -----
BRA1 Sequence 4 -----
BRA2 sequence 3.1 -----
BRA1 Sequence 2 -----

BRA1
BRA1 Sequence 3 TTATTTTAGTTACATGACTAACAGACGTATACATGTACAAAATAACACTGTTTCA-TTAT
BRA2 sequence 4.1 TTATTTTAGTTACATGACTAACAGACGTATACATGTACAAAATAACACTGTTTCMWTAT
-----
BRA1 Sequence 1 TATATTAGCTTCAGTAACCAGCAGAAGTAGACATGTACAAAATAACACTGTTTCA-TTAT
BRA2 sequence 5.1 -----
BRA1 Sequence 2 -----CAGTAGTAGACATGTACAAAATAACACTGTTTCA-TTAT
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 -----
BRA2 sequence 4.2 -----
BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 -----
BRA1 Sequence 4 -----
BRA2 sequence 3.1 -----
BRA1 Sequence 2 -----

BRA1
BRA1 Sequence 3 TAG-AGTCATATCATAGAGAGATAACATATATATATAAAA-CAATGTGTACAGATAGTGA
BRA2 sequence 4.1 TAG-AGTCATATCATAGAGAGATAACATATATATATAAAAACAATGTGTACAGATAGTGA
-----
BRA1 Sequence 1 TAG-AGTCATATCGTAGAGAGATAGCATTATATATATACAA-CAATGTCTACTGGTAGTGA
BRA2 sequence 5.1 -----
BRA1 Sequence 2 CAGGAGTCATATCATAGAGAGATAACATATATATATACAA-CAATGTGTACTGATAGTGA
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 -----TGTGTACAGATAGTGA
BRA2 sequence 4.2 -----TGTGTACAGATAGTGA
BRA2 sequence 5.2 -----TGTGTACAGATAGTGA
BRA2 sequence 3.2 -----TGTGTACAGATAGTGA
BRA2 sequence 6.2 -----GTGTACAGATAGTGA
BRA1 Sequence 1 -----AGTGA
BRA2 sequence 1.1 -----
BRA1 Sequence 4 -----
BRA2 sequence 3.1 -----
BRA1 Sequence 2 -----

BRA1
BRA1 Sequence 3 TCATTGTCAAGAA-TAT-GAGTGT-----AGAAACG--TACAACTGACT-
BRA2 sequence 4.1 TCATTGTCAAGAANTATNGAGTGT-----YMAAGAGGGTACAACTGACTA
-----
BRA1 Sequence 1 TCATCGTCAAGAT-TAT-GAGKWT-----RAARAGG--TACAACTGACT-
BRA2 sequence 5.1 -----
BRA1 Sequence 2 TCATTATCAAGAT-TAT-GAGTGT-----GAGGAGG--TACAACTGACT-
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 TCATTGTCAAGAC-TAT-GAGTGC GAAAGGGTCCAACATAAAAAACA-CTAGTAAATATA-
BRA2 sequence 4.2 TCATTGTCAAGAC-TAT-GAGTGC-----GAAAGGG--TCCAACATAAAT-
BRA2 sequence 5.2 TCATTGTCAAGAC-TAT-GAGTGC-----AAAAGGG--TCCAACATAAAA-
BRA2 sequence 3.2 TCATTGTCAASAC-TAT-GAGTGC-----GAAAGGG--TCCGATTAAAT-
BRA2 sequence 6.2 TCATTGTCAAGAC-TAT-GAGTGC-----AAAAGGG--TCCAACATAAAA-
BRA1 Sequence 1 TCATTGTCAAGAA-TAT-GAGTGT-----AAAAAGG--AACAACTGCCT-
BRA2 sequence 1.1 -----
BRA1 Sequence 4 -----
BRA2 sequence 3.1 -----
BRA1 Sequence 2 -----

```

```

BRA1 G----G-----TAAT----GGGTAAAA-GTG--AATATAA-GCTACACCTTACTA
BRA1 Sequence 3 G----G-----TATR----TGGGTCNAA-GTGTGACTATAAGCTACAACG--CTA
BRA2 sequence 4.1 -----
BRA1 Sequence 1 A----G-----TART----GGGTCAAGT-GTG--ACTATAA-GCTACAAC--RCTA
BRA2 sequence 5.1 -----
BRA1 Sequence 2 G----G-----TGAG----TAGTTGTAGTGTG--AATATAA-CCTACAAC--ACTA
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 T----A-----ATAA----ATGTGACGG-TTA--GTAATAA-TAGTGGTAATAGTG
BRA2 sequence 4.2 GAGAA-----TAAT----AGTGGAAGT-GTG--GAGATTA-GTGGTAA--TAGCA
BRA2 sequence 5.2 G----A GACTAAAA TAAT----ATAATAAAT-GTG--AGTAATA-GTGGTAA--TAGTA
BRA2 sequence 3.2 G----G-----GAATTTTAGTGATAGAA-TAA--GTGTGGA-GGATAATAATAATA
BRA2 sequence 6.2 G----A GACTAAAA TAAT----ATAATAAAT-GTG--AGTAATA-GTGGTAA--TAGTA
BRA1 Sequence 1 A----G-----TAAT----GGGTCCCGT-GTG--AATATAA-GCTACAACG--CTA
BRA2 sequence 1.1 -----
BRA1 Sequence 4 -----
BRA2 sequence 3.1 -----
BRA1 Sequence 2 -----

```

```

BRA1 ACACATC--AT----TGGAA-TTTCAAATGGTT--GGTATGTAATCAAAGTGAAGATCAA
BRA1 Sequence 3 GCACATC--TTAATTTGGAANTTTCAAATGGGTTGGGTATGTAATCAAAGTGRAGAATCAA
BRA2 sequence 4.1 -----
BRA1 Sequence 1 GCACATCTTAT----TGGAA-TTTCAAATGGTT--GGTATGTAATCAAAGTGAAGATCAA
BRA2 sequence 5.1 -----
BRA1 Sequence 2 MCACATCTTAT----TGGAA-TTTCAAATGGTT--GGTATGTAATCAAAGTGAAGATCAA
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 ACACTT---AT----TGGGA-TTTCACATGGTT--ATTATGTAATCAAAGTGAAGATCAA
BRA2 sequence 4.2 GCACTT---AT----TGGGA-TTTCACATGGTT--ATTATGTAATCAAAGTGAAGATCAA
BRA2 sequence 5.2 GCACTT---AT----TGGGA-TTTCACATGGTT--ATTATGTAATCAAAGTGAAGATCAA
BRA2 sequence 3.2 GCACTT---AT----TGGGA-TTTCACATGGTT--ATTATGTAATCAAAGTGAAGATCAA
BRA2 sequence 6.2 GCACTT---AT----TGGGA-TTTCACATGGTT--ATTATGTAATCAAAGTGAAGATCAA
BRA1 Sequence 1 GCACATCTTAT----TGGAA-TTTCAAATGGTT--GGTATGTAATCAAAGTGAAGATCAA
BRA2 sequence 1.1 -----
BRA1 Sequence 4 -----
BRA2 sequence 3.1 -----
BRA1 Sequence 2 -----

```

```

BRA1 AAGGCAATTC-TTATTCCA-GAAGAAG-AAATGGTACAG---AA-TGATAATAATACTT-
BRA1 Sequence 3 AAGGCAATTCYTATTCCAAGAAGAAGNAAATGGTACAG---AAATGATAATAATACYT-
BRA2 sequence 4.1 -----
BRA1 Sequence 1 AAGGCAATTC-TTATTCCA-GAAGAAG-AAATGGTACAG---AA-TGATAATAATACTT-
BRA2 sequence 5.1 -----
BRA1 Sequence 2 AAGGCAATTC-TTATTSCA-GAAGAAG-AAGTGGTACAG---AA-TGATAATAATACTT-
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 AAGACAATTC-TTGTTCCTA-GAAGAAG-AAATGATACTG---TA-CAATGGCACCCTT-
BRA2 sequence 4.2 AAGACAATTC-TTGTTCCTA-GAAGAAG-AAATGATACAGTACAA-TGACAATCGCACTT-
BRA2 sequence 5.2 AAGACAATTC-TTGTTCCTA-GAGGAAG-AAATGCAGGAC---AG-TGACAATGGCACTT-
BRA2 sequence 3.2 AAGACAATTC-TTGTTCCTA-GAAGAAG-AAATGATACAGTACAA-TGACAATCGCACTT-
BRA2 sequence 6.2 AAGACAATTC-TTGTTCCTA-GAGGAAG-AAATGCAGGAC---AG-TGACAATGGCACTT-
BRA1 Sequence 1 AAGGCAATTC-TTATTCCA-GAAGAAG-AAATGGTACAG---AA-TGATAATAATACTT-
BRA2 sequence 1.1 -----
BRA1 Sequence 4 -----AATAATACTT-
BRA2 sequence 3.1 -----
BRA1 Sequence 2 -----AATAATACTT-

```

```

BRA1 -GGTCCCCTAGGGGAT-GTAATAGCA--CATGGG-CAAGGGTGAAACATT-GTCC-AGCG
BRA1 Sequence 3 GGGTCCCCTAGGGGATTGTAATAGCCACCATGGGGCAAAGGGGAAACACTTGTCAGGCCA
BRA2 sequence 4.1 -----
BRA1 Sequence 1 -GGTCCCCTAGGGGAT-GTAATAGCA--CATGGG-CAAAGGTGAAACATT-GTCC-AGCG
BRA2 sequence 5.1 -----
BRA1 Sequence 2 -GGTCTCGTAGGGGAT-GTARTAGCA--CATGGR-CAAAGGTGAAACATT-GTCC-AGCT
BRA2 sequence 2.1 -----

```

```

BRA2 sequence2.2 -GGGCCCCTAAGGGAT-GTAATGACA--CATGGG-CAAAAGTAAAACATT-GTCC-AATG
BRA2 sequence4.2 -GGGCCCCTAAGGGAT-GTAATGACA--CATGGG-CAAAAGTAAAACATT-GTCC-AATG
BRA2 sequence5.2 -GGGCCCCTAAGGGAT-GTAATGACA--CATGGG-CAAAAGTAAAACATT-GTCC-AATG
BRA2 sequence3.2 -GGGCCCCTAAGGGAT-GTAATGACA--CATGGG-CAAAAGTAAAACATT-GTCC-AATG
BRA2 sequence6.2 -GGGCCCCTAAGGGAT-GTAATGACA--CATGGG-CAAAAGTAAAACATT-GTCC-AATG
BRA1 Sequence 1 NGGTCCCCTAGGGGAT-GTAATAGCA--CATGGG-CAAAGGTGAAACATT-GTCC-AGCG
BRA2 sequence1.1 -----
BRA1 Sequence 4 -GGTCCCCTAGGGGAT-GTAATAGCA--CATGGG-CAAAGGTGAAACATTTGTCCCAGCG
BRA2 sequence3.1 -----
BRA1 Sequence 2 -GGTCCCCTAGGGGAT-GTAATAGCA--CATGGG-CAAAGGTGAAACATTTGTCCCAGCG

```

```

BRA1 GAT-CTGTTGTATGGGATACATCCTATTAGGCTATGTGTGCAGCCACCATTTTCTGGT
BRA1 Sequence 3 GATCGGRWTCATGCGATACATCCTATAAGGCCTATGTTGTGCAGCCACATTTTCTGGT
BRA2 sequence 4.1 -----
BRA1 Sequence 1 GAT-YTGTTGTATGGGATACATCCTATTAGGCTATGTGTGCAGCCACCATTTTCTGGT
BRA2 sequence 5.1 -----
BRA1 Sequence 2 GAT-STGTTKTATGGGATTCATCYTATTAGGCTATGTGTGCAGCCACCATTTTCTGGT
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 GAC-TTGTTGTATGGGCTGCATCCTATTAGGTTATGTGTACAGCCACCATTTTCTGGT
BRA2 sequence 4.2 GAC-TTGTTGTATGGGCTGCATCCTATTAGGTTATGTGTACAGCCACCATTTTCTGGT
BRA2 sequence 5.2 GAC-TTGTTGTATGGGCTGCATCCTATTAGGTTATGTGTACAGCCACCATTTTCTGGT
BRA2 sequence 3.2 GAC-TTGTTGTATGGGCTGCATCCTATTAGGTTATGTGTACAGCCACCATTTTCTGGT
BRA2 sequence 6.2 GAC-TTGTTGTATGGGCTGCATCCTATTAGGTTATGTGTACAGCCACCATTTTCTGGT
BRA1 Sequence 1 GATTYTGTTGTATGGGATACATCCTATTAGGCTATGTGTGCAGCCACCATTTTCTGGT
BRA2 sequence 1.1 -----
BRA1 Sequence 4 GAT-TTGTTGTATGGGATACATCCTATTAGGCTATGTGTGCAGCCACCATTTTCTGGT
BRA2 sequence 3.1 -----
BRA1 Sequence 2 GAT-TTGTTGTATGGGATACATCCTATTAGGCTATGTGTGCAGCCACCATTTTCTGGT

```

```

BRA1 AAAAAATAATGATGGGAACCTCACAAATGGGAACCCACACACTTAGCAACTGTGGTCCACA
BRA1 Sequence 3 AAAATATGGATTGGAACCTCAGAAATATGGGGAACCTCAACCACTCCTA-----
BRA2 sequence 4.1 -----
BRA1 Sequence 1 RAAAAATAATGATGGGAACCTCACAAATGGGAACCCACACACTTAGCAACTGTGGTCCACA
BRA2 sequence 5.1 -----
BRA1 Sequence 2 AAAAAATAATGATGGGAACCTCACATTTGGGAACCCACACACTYAGCAAYTGTGGTCCACA
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 AAGAAAAGAAGAT-----AATCATAGCCGCACACTTAGTAATTGTGGTCCACA
BRA2 sequence 4.2 AAGAAAAGAAGAT-----AATCATAGCCGCACACTTAGTAATTGTGGTCCACA
BRA2 sequence 5.2 AAGAAAAGAAGAT-----AATCATAGCCGCACACTTAGTAATTGTGGTCCACA
BRA2 sequence 3.2 AAGAAAAGAAGAT-----AATCATAGCCGCACACTTAGTAATTGTGGTCCACA
BRA2 sequence 6.2 AAGAAAAGAAGAT-----AATCATAGCCGCACACTTAGTAATTGTGGTCCACA
BRA1 Sequence 1 GAAAAATAATGATGGGAACCTCACAAATGGGAACCCACACACTTAGCAACTGTGGTCCACA
BRA2 sequence 1.1 -----
BRA1 Sequence 4 GAAAAATAATGATGGGAACCTCACAAATGGGAACCCACACACTTAGCAACTGTGGTCCACA
BRA2 sequence 3.1 -----
BRA1 Sequence 2 GAAAAATAATGATGGGAACCTCACAAATGGGAACCCACACACTTAGCAACTGTGGTCCACA

```

```

BRA1 AATATCACTGGGGATATTAGATGAAAATAAAGCTGTTACAAGGGAGGTCAACTGCACTGT
BRA1 Sequence 3 -----
BRA2 sequence 4.1 -----
BRA1 Sequence 1 AATATCACTGGGGATATTAGATGAAAATAAAGCTGTTACAAGGGGGGTCAACTGCACTGT
BRA2 sequence 5.1 -----
BRA1 Sequence 2 AAYAYMACTGGGGATATTRGATGAAAATAAAGCTGTTACAAGGGGGGTCAACTCACTGT
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 AATATCATTAGGGATATTAGATAACAATAAAGCTGCTGTAAGAGATGCCAGCTGCCAGGT
BRA2 sequence 4.2 AATATCATTAGGGATATTAGATGACAATAAAGCTGCTGTAAGAGATGCCAAGTGCAGGT
BRA2 sequence 5.2 AATATCACTAGGGATATTAGATAACAATAAAGCTGCTGTAAGAAATGCCAAGTGCAGCGT
BRA2 sequence 3.2 AATATCATTAGGGATATTAGATAGCAATAAAGCTGCTGTAAGAAATGCCAAGTGCAGGT
BRA2 sequence 6.2 AATATCACTAGGGATATTAGATAACAATAAAGCTGCTGTAAGAAATGCCAAGTGCAGCGT
BRA1 Sequence 1 AATATCACTGGGGATATTAGATGAAAATAAAGCTGTTACAAGGGGGGTCAACTGCACTGT
BRA2 sequence 1.1 -----
BRA1 Sequence 4 AATATCACTGGGGATATTAGATGAAAATAAAGCTGTTATAAGGGAGGTCAACTGCACTGT

```

```

BRA2 sequence 3.1 -----
BRA1 Sequence 2 AATATCACTGGGGATATTAGATGAAAATAAAGCTGTTATAAGGGAGGTCAACTGCACTGT

BRA1
GGCTAAAAAA-GTTTTTCATTTT---AAT---GACTATTCTGGACAGTTTATAAGTCCTA
BRA1 Sequence 3 -----
BRA2 sequence 4.1 -----
BRA1 Sequence 1 GGCTAAAAAA-AGTTTTAATTTT---CAT---GACTATTCTGGACAGTTTATAAGYCCTA
BRA2 sequence 5.1 -----
BRA1 Sequence 2 GGCTAAAAAA-AGTTATAATTTT---CAT---GACTATTCTGGACAGWTTATAAGTCTAA
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 GGAGAGAAAAG-ACTGTTTCAGATT---CCTAAAGAATATTCTGGACAGCTTATAGTTCCTA
BRA2 sequence 4.2 GGAAAAAAAG-ACTGTTTCAGATT---CCTAAAGAATATTCTGGACAGCTTATAGTTCCTA
BRA2 sequence 5.2 AGAGAAAAAG-ACTGTTTCAGACT---AAA---GACTATTCTGGACAGCTTATAGTTCCTA
BRA2 sequence 3.2 GGAGAAAAAG-ACTTTTTCAGTTTCTAAA---GGCCATTCTGGACAGCTTATAGTTCCTA
BRA2 sequence 6.2 AGAGAAAAAG-ACTGTTTCAGACT---AAA---GACTATTCTGGACAGCTTATAGTTCCTA
BRA1 Sequence 1 GGCTAAAAAAAGTTTTTNAATTTT---CAT---GACTATTCTGGACAGTTTATAAGCCCTA
BRA2 sequence 1.1 -----
BRA1 Sequence 4 GGCTAAAAAA-AGTTTTAATTTT---CAT---GACTATTCTGGACAGTTTATAAGTCCTA
BRA2 sequence 3.1 -----
BRA1 Sequence 2 GGCTAAAAAA-AGTTTTAATTTT---CAT---GACTATTCTGGACAGTTTATAAGTCCTA

BRA1
TATTTTATCAATGTAATCTTAGC---CTAACACCTTGTAATGGGACAAACTCACC-----
BRA1 Sequence 3 -----
BRA2 sequence 4.1 -----
BRA1 sequence 1 TATTTTATCAATGTAATCTTAGC---CTAAACCTTGTAATGGGACAAACTCATC-----
BRA2 sequence 5.1 -----
BRA1 Sequence 2 TATTTTATCAATGTAATCTTAGY---CTAAAMCCTTGTAATGGGACAAACTCATC-----
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 TATTCTATAAGTGTAATCTTAGC---CTAATCTGACAAGGTGCTATAACAACAGCTCAAC
BRA2 sequence 4.2 TATTCTATAAGTGTAATCTTAGC---CTAATCTGACAAGGTGCAATAACGACAACTCAAC
BRA2 sequence 5.2 TATTCTATAAGTGTAATCTTAGC---CTAATCTGACAAGGTGCGATAACAACAACCTCAAC-----
BRA2 sequence 3.2 TATTCTATAAGTGTAATCTTAGC---CTAATCTGACAAGGTGCTATAACAACAGCTCAAC-----
BRA2 sequence 6.2 TATTCTATAAGTGTAATCTTAGC---CTAATCTGACAAGGTGCGATAACAACAACCTCAAC-----
BRA1 Sequence 1 TATTTTATCAATGTAATCTTAGC---CTAAACCTTGTAATGGGACAAACTCATC-----
BRA2 sequence 1.1 -----
BRA1 Sequence 4 TATTTTATCAATGTAATCTTAGC---CTAAACCTTGTAATGGGACAAACTCATC-----
BRA2 sequence 3.1 -----
BRA1 Sequence 2 TATTTTATCAATGTAATCTTAGC---CTAAACCTTGTAATGGGACAAACTCATC-----

BRA1
AGTAAGTATTATCAGATATGATGAGGACACTGTTGAATATTTGATAAGAAACACTGTTGA
BRA1 Sequence 3 -----
BRA2 sequence 4.1 -----
BRA1 Sequence 1 AGTAAGTATTATCAGATATGATGAG-----AACACTGTTGA
BRA2 sequence 5.1 -----
BRA1 Sequence 2 AGTAAGTATTATCAGATATGATGAGGACGCTGTYGAATATGTGATA-----
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 AATAAGTATTATCAGGTATGATGAGAAAAATGTTCAATATTTGTTATGTAAAACAAATAA
BRA2 sequence 4.2 AATAAGTATTATCAGATATGATAAGGACAATGTTCAATATTTGTTATGTAAAATAATAC
BRA2 sequence 5.2 AATAAGTGTTATCAGGTATGATGAGAACAGTGTTCAATATTTGTTATGTAAAACAAATAA
BRA2 sequence 3.2 AATAAGTATTATCAGGTATGATGAG-----AAAA-TGTTCA
BRA2 sequence 6.2 AATAAGTGTTATCAGGTATGATGAGAACAGTGTTCAATATTTGTTATGTAAAACAAATAA
BRA1 Sequence 1 AGTAAGTATTATCAGATATGATGAG-----AA-CAGTGTGA
BRA2 sequence 1.1 -----
BRA1 Sequence 4 AGTAAGTATTATCAGATATGATGAGAACTGTTGAGTATTTGATA-----
BRA2 sequence 3.1 -----
BRA1 Sequence 2 AGTAAGTATTATCAGATATGATGAGAACTGTTGAGTATTTGATA-----

BRA1
ATATTTGATATGCAAGGCTAAAAACATCACTGAAAA-----A-----
BRA1 Sequence 3 -----
BRA2 sequence 4.1 -----
BRA1 Sequence 1 ATATTTGATATGCAAGGCTAATAGCACCCTGAAAAACACCCTGAAGGA-----

```

```

BRA2 sequence 5.1 -----
BRA1 Sequence 2 -----TGCAAGGMTAATAGCAACACTGAAAACACCACTGAAGGA-----
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 TACAGATGCAGGTAAT-----
BRA2 sequence 4.2 AGATGCAGGT-----
BRA2 sequence 5.2 TACAGATGCAGGTAAT-----
BRA2 sequence 3.2 ATATTTGTTATGTAAAACAAATAATACAGATGCAGG-----T-----
BRA2 sequence 6.2 TACAGATGCAGGTAAT-----
BRA1 Sequence 1 ATATTTGATATGCAAGGCTAATAGCACCACTGAAAACACCACTGAAGGA-----
BRA2 sequence 1.1 -----
BRA1 Sequence 4 -----TGCAAGGCTAATAACATCACT-AAAA-----ACACCGCTGAAA
BRA2 sequence 3.1 -----
BRA1 Sequence 2 -----TGCAAGGCTAATAACATCACT-AAAA-----ACACCGCTGAAA

```

```

BRA1 -----AACTTTTCATGTGTAGTACAACTTTTGGAAACAATAGGACAGGCAC
BRA1 Sequence 3 -----
BRA2 sequence 4.1 -----
BRA1 Sequence 1 -----AGCTTTTCATGTGTAGTACAACTTTTGGAAACAATAGGACAGGCAC
BRA2 sequence 5.1 -----
BRA1 Sequence 2 -----AACTTTTCATGTGTAGTACAACTTTTGGAAACAATAGGACAGGCAC
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 -----TTTTCATGTGTAGTACAACTTTTGGGAAAATAGGACAGGCAC
BRA2 sequence 4.2 -----AATTTTCATGTGTAGTACAACTTTTGGGAAAATAGGACAGGCAC
BRA2 sequence 5.2 -----TTTTCATGTGTAGTACAACTTTTGGGAAAATAGGACAGGCAC
BRA2 sequence 3.2 -----AATTTTCATGTGTAGTACAACTTTTGGGAAAATAGGACAGGCAC
BRA2 sequence 6.2 -----TTTTCATGTGTAGTACAACTTTTGGGAAAATAGGACAGGCAC
BRA1 Sequence 1 -----AGCTTTTCATGTGTAGTACAACTTTTGGAAACAATAGGACAGGCAC
BRA2 sequence 1.1 -----
BRA1 Sequence 4 GCACCACTGAAGGAACTTTTCATGTGTAGTACAACTTTTGGAAACAATAGGACAGGCAC
BRA2 sequence 3.1 -----
BRA1 Sequence 2 GCACCACTGAAGGAACTTTTCATGTGTAGTACAACTTTTGGAAACAATAGGACAGGCAC

```

```

BRA1 ATATAGAACTACCCAGAAAAATAAAAAGATAATTGGAAACCAGTTCACACAATATAATT
BRA1 Sequence 3 -----
BRA2 sequence 4.1 -----
BRA1 Sequence 1 ATATAGAACTACCCAGAAAAATAAAAAGATAATTGGAGACCGGTTACACAATATAATT
BRA2 sequence 5.1 -----
BRA1 Sequence 2 ATATAGAACTACCCAGAAAAATAAAAAGATAATTGGAAACCAGTTCACACAATATGATT
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 ATATAGAGTTACCCAGAAACAATAAAAGGATAAGAGAGAACAATTCACACAGTATAATT
BRA2 sequence 4.2 ATATAGAGTTACCCAGAAACAATAAAAGGATAAGAGAGAACAATTCACACAGTATAATT
BRA2 sequence 5.2 ATATAGAGTTACCCAGAAACAATAAAAGGATAAGAGAGAACAATTCACACAGTATAATT
BRA2 sequence 3.2 ATATAGAGTTACCCAGAAACAATAAAAGGATAAGAGAGAACAATTCACACAGTATAATT
BRA2 sequence 6.2 ATATAGAGTTACCCAGAAACAATAAAAGGATAAGAGAGAACAATTCACACAGTATAATT
BRA1 Sequence 1 ATATAGAACTACCCAGAAAAATAAAAAGATAATTGGAGACCGGTTACACAATATAATT
BRA2 sequence 1.1 -----
BRA1 Sequence 4 ATATAGAACTACCCAGAAAAATAAAAAGATAATTGGAGACCGGTTACACAATATAATT
BRA2 sequence 3.1 -----
BRA1 Sequence 2 ATATAGAACTACCCAGAAAAATAAAAAGATAATTGGAGACCGGTTACACAATATAATT

```

```

BRA1 GTTCAATAAATGATAAGACAGAGTTAGAAAAATGGAAGTTGGTAAAAA-GTTCAGGTATT
BRA1 Sequence 3 -----
BRA2 sequence 4.1 -----
BRA1 Sequence 1 GTTCAATAAATAATAAGACAGAGTTAGAAAAATGGAAGTTGGTAAAAA-GTTCAGGTATT
BRA2 sequence 5.1 -----
BRA1 Sequence 2 GTTCAATAAATAATAAGACAGAGTTAGAAAAATGGAAGTN-----
BRA2 sequence 2.1 -----
BRA2 sequence 2.2 GTTCAATAAATAATCAGACAGAGTTAAAAGAATGGAAATTGATAAAG-GTTCAGGTATT
BRA2 sequence 4.2 GTTCAATAAATAATCAGACAGAGTTAAAAGAATGGAAATTGATAAAG-GTTCAGGTATT
BRA2 sequence 5.2 GTTCAATAAATAATCAGACAGAGTTAAAAGAATGGAAATTGATAAAG-GTTCAGGTATT
BRA2 sequence 3.2 GTTCAATAAATAATCAGACAGAGTTAAAAGAATGGAAATTGATAAAG-GTTCAGGTATT
BRA2 sequence 6.2 GTTCAATAAATAATCAGACAGAGTTAAAAGAATGGAAATTGATAAAG-GTTCAGGTATT

```

```

BRA1 Sequence 1      GTTCAATAAATAATAAGACAGAGTTAGAAAAATGGAAGTTGGTAAAAANGTTCAGGTATT
BRA2 sequence 1.1    -----
BRA1 Sequence 4      GTTCAATAAATGATAAGACAGAGTTAGAAAAATGGAAGTTGGTAAAAA-GTTCAGGTATT
BRA2 sequence 3.1    -----
BRA1 Sequence 2      GTTCAATAAATGATAAGACAGAGTTAGAAAAATGGAAGTTGGTAAAAA-GTTCAGGTATT

```

```

BRA1                  ACTCCTTTTCCTGTTACTTCTCGGGCCAATACAGGACTAGTTAGATACAAAAGAGATTTT
BRA1 sequence 3      -----
BRA2 sequence 4.1    -----
BRA1 Sequence 1      ACTCCTTTTCCTGTTACTTCTCGGGC-----
BRA2 sequence 5.1    -----
BRA1 Sequence 2      -----
BRA2 sequence 2.1    -----
BRA2 sequence 2.2    ACTCCTATTTCCTATTACTTCTCAGGCTAACACAGGATTAATTAGATACAAAAGAGATTTT
BRA2 sequence 4.2    ACTCCTGTTTCCTATTACTTCTCAGGCTAACACAGGATTAATTAGATACAAAAGAGATTTT
BRA2 sequence 5.2    ACTCCTATTTCCTATTACTTCTCAGGCTAACACAGGATTAATTAGATACAAAAGAGATTTT
BRA2 sequence 3.2    ACTCCTATTTCCTATTACTTCTCAGGCTAACACAGGATTAATTAGATACAAAAGAGATTTT
BRA2 sequence 6.2    ACTCCTATTTCCTATTACTTCTCAGGCTAACACAGGATTAATTAGATACAAAAGAGATTTT
BRA1 Sequence 1      ACTCCTTTTCCTGTTACTTCTCGGGC-----
BRA2 sequence 1.1    -----
BRA1 Sequence 4      ACTCCTTTTCCTGTTACTTCTCGGGC-----
BRA2 sequence 3.1    -----
BRA1 Sequence 2      ACTCCTTTTCCTGTTACTTCTCGGGC-----

```

```

BRA1                  GGAATATCAGCAATAGTGGCGGCAATAGTGGCTGCAACTGCTATTGCTGCTAGTGCTACT
BRA1 Sequence 3      -----
BRA2 sequence 4.1    -----GCTATTGCTGCTAGTGCTACC
BRA1 Sequence 1      -----
BRA2 sequence 5.1    -----
BRA1 Sequence 2      -----
BRA2 sequence 2.1    -----
BRA2 sequence 2.2    GGAATATCAGCAATAGTGGCGGCAATAGTGGCTGCAACCGCTATTGCTGCTAGTGCTACC
BRA2 sequence 4.2    GGAATATCAGCAATAGTGGCGGCAATAGTGGCTGCAACCGCTATTGCTGCTAGTGCTACC
BRA2 sequence 5.2    GGAATATCAGCAATAGTGGCGGCAATAGTGGCTGCAACCGCTATTGCTGCTAGTGCTACC
BRA2 sequence 3.2    GGAATATCAGCAATAGTGGCGGCAATAGTGGCTGCAACCGCTATTGCTGCTAGTGCTACC
BRA2 sequence 6.2    GGAATATCAGCAATAGTGGCGGCAATAGTGGCTGCAACCGCTATTGCTGCTAGTGCTACC
BRA1 Sequence 1      -----
BRA2 sequence 1.1    -----
BRA1 Sequence 4      -----
BRA2 sequence 3.1    -----
BRA1 Sequence 2      -----

```

```

BRA1                  ATGTCTTATATTGCCTTAACTGAAACAAATAAAATGAGTGATGTAATGAATCATACATTT
BRA1 Sequence 3      -----
BRA2 sequence 4.1    ATGTCTTATATTGCCTTGACTGAAGCAAGTAAAGTTACTAATGCCATGAATCATACATTT
BRA1 Sequence 1      -----
BRA2 sequence 5.1    -----TAATGCCATGAATCATACMTTT
BRA1 Sequence 2      -----
BRA2 sequence 2.1    -----TTT
BRA2 sequence 2.2    ATGTCTTATATTGCTTTGACTGAAGCAAGTAAAGTTACTAATGCCATGAATCATACCTTT
BRA2 sequence 4.2    ATGTCTTATATTGCTTTGACTGAAGCAAGTAAAGTTACTAATGCCATGAATCATACCTTT
BRA2 sequence 5.2    ATGTCTTATATTGCTTTGACTGAGGCAAGTAAAGTTAGTAATGCCATGAATCATACATTT
BRA2 sequence 3.2    ATGTCTTATATTGCTTTGACTGAAGCAAGTAAAGTTAGTAATGCCATGAATCATACATTT
BRA2 sequence 6.2    ATGTCTTATATTGCTTTGACTGAGGCAAGTAAAGTTAGTAATGCCATGAATCATACATTT
BRA1 Sequence 1      -----
BRA2 sequence 1.1    -----
BRA1 Sequence 4      -----
BRA2 sequence 3.1    -----
BRA1 Sequence 2      -----

```

```

BRA1                  GTAGTGAAAAATGATACAATCAGAGCAACGGAAATGATAGAAAAACAGATACATATATTA

```

```

BRA1 Sequence 3 -----
BRA2 sequence 4.1 GAAGTAGAGAATAACACAATCCATGCAACAGAGATGATAGAAAAACAAATGCATATATTA
BRA1 Sequence 1 -----
BRA2 sequence 5.1 GAAGTAGAGAATAACACAATCCATGCAACAGAGATGATAGAAAAACAAATGCATATATTA
BRA1 Sequence 2 -----
BRA2 sequence 2.1 GAAGTAGAGAATAACACAATCCATGCAACAGAGATGATAGAAAAACAAATGCATATATTA
BRA2 sequence 2.2 GAAGTAGAGAATAACACAATCCATGCAACAGAGATGATAGAAAAACAAATGCATATATTA
BRA2 sequence 4.2 GAAGTAGAGAATAACACAATCCATGCAACAGAGATGATAGAAAAACAAATGCATATATTA
BRA2 sequence 5.2 GAGGTAGAGAATAACACAATCCATGCAACAGAGATGATAGAAAAACAAATGCATATATTA
BRA2 sequence 3.2 GAGGTAGAGAATAACACAATCCATGCAACAGAGATGATAGAAAAACAAATGCATATATTA
BRA2 sequence 6.2 GAGGTAGAGAATAACACAATCCATGCAACAGAGATGATAGAAAAACAAATGCATATATTA
BRA1 Sequence 1 -----
BRA2 sequence 1.1 ---GTAGAGAATAACACAATCCATGCAACAGAGATGATAGAAAAACAAATGCATATATTA
BRA1 Sequence 4 -----
BRA2 sequence 3.1 -----AAAAACAAATGCATATATTA
BRA1 Sequence 2 -----

```

```

BRA1 TATGCAATGGTTCTTCAGACCCATGCAGATGTCC-AGTTACTAAAAGAAAGACAAAAAAT
BRA1 Sequence 3 -----
BRA2 sequence 4.1 TATGCAATGATTCTTCAGACCCATGCAGATGTCC-AGTTACTAAAAGAAAGACAAAAAAT
BRA1 Sequence 1 -----
BRA2 sequence 5.1 TATGCAATGATTCTTCAGACCCATGCAGATGTCC-AGTTACTAAAAGAAAGACAAAAAAT
BRA1 Sequence 2 -----
BRA2 sequence 2.1 TATGCAATGATTCTTCAGACCCATGCAGATGTCC-AGTTACTAAAAGAAAGACAAAAAAT
BRA2 sequence 2.2 TATGCAATGATTCTTCAGACCCATGCAGATGTCC-AGTTACTAAA-----
BRA2 sequence 4.2 TATGCAATGATTCTTCAGACCCATGCAGATGTCC-AGTTACTAAA-----
BRA2 sequence 5.2 TATGCAATGATTCTTCAGACCCATGCAGATGTCC-AGTTA-----
BRA2 sequence 3.2 TATGCAATGATTCTTCAGACCCATGCAGATGTCC-AGTTACTAAA-----
BRA2 sequence 6.2 TATGCAATGATTCTTCAGACCCATGCAGATGTCC-AG-----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 TATGCAATGATTCTTCAGACCCATGCAGATGTCC-AGTTACTAAAAGAAAGACAAAAAAT
BRA1 Sequence 4 -----
BRA2 sequence 3.1 TATGCAATGATTCTTCAGACCCATGCAGATGTCCAGTTACTAAAAGAAAGACAAAAAAT
BRA1 Sequence 2 -----

```

```

BRA1 AGAGGAAACATTTAATTTG-ATAGGGTGTGTTGAGAGGTCACATACATTCTGCCATACTG
BRA1 Sequence 3 -----
BRA2 sequence 4.1 AGAGGAAACATTTAATTTG-ATAGGGTGTGTTGAAAGATCACATACATTCTGCCATACTG
BRA1 Sequence 1 -----
BRA2 sequence 5.1 AGATGAAACATTTAATTTG-ATAGGGTGTGTTGAAAGATCACATACATTCTGTCATACTG
BRA1 Sequence 2 -----
BRA2 sequence 2.1 AGATGAAACATTTAATTTG-ATAGGGTGTGTTGAAAGATCACATACATTCTGTCATACTG
BRA2 sequence 2.2 -----
BRA2 sequence 4.2 -----
BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 AGATGAAACATTTAATTTG-ATAGGGTGTGTTGAAAGATCACATACATTCTGTCATACTG
BRA1 Sequence 4 -----
BRA2 sequence 3.1 AGATGAAACATTTAATTTG-ATAGGGTGTGTTGAAAGATCACATACATTCTGTCATACTG
BRA1 Sequence 2 -----

```

```

BRA1 GACA-TCCATGGAATGAATCTTGGGGGCAAT-TGAATGAAACAACAGAATGGGA
BRA1 Sequence 3 -----
BRA2 sequence 4.1 GACA-TCCATGGAATGATTCTTGGGGACAAT-TAAATGACTCAACACAATGGGA
BRA1 Sequence 1 -----
BRA2 sequence 5.1 GACA-TCCATGGAATGATTCTTGGGGACAAT-TAAATGACTCAACACAATGGGA
BRA1 Sequence 2 -----
BRA2 sequence 2.1 GACA-TCCATGGAATGATTCTTGGGGACAAT-TAAATGACTCAACACAATGGGA
BRA2 sequence 2.2 -----
BRA2 sequence 4.2 -----

```

```

BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 GACA-TCCATGGAATGATTCTTGGGGACAAT-TAAATGACTCAACACAATGGGA
BRA1 Sequence 4 -----
BRA2 sequence 3.1 GACA-TCCATGGAATGATTCTTGGGGACAAT-TAAATGACTCAACACAATGGGA
BRA1 Sequence 2 -----

BRA1 TGATTGGATAACTAAAATGAATAGGCTTGAAACGGAGATATTGACTACTCTGCATGCAGC
BRA1 Sequence 3 -----
BRA2 sequence 4.1 TGATTGGGTAAACAAAATGAATGAGTATGAACAGGACATATTAACAACCTCTTCATGCAGC
BRA1 Sequence 1 -----
BRA2 sequence 5.1 TGATTGGGTAAACAAAATGAATGAGTATGAACAGGACATATTAACAACCTCTTCATGCAGC
BRA1 Sequence 2 -----
BRA2 sequence 2.1 TGATTGGGTAAACAAAATGAATGAGTATGAACAGGACATATTAACAACCTCTTCATGCAGC
BRA2 sequence 2.2 -----
BRA2 sequence 4.2 -----
BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 TGATTGGGTAAACAAAATGAATGAGTATGAACAGGACATATTAACAACCTCTTCATGCAGC
BRA1 Sequence 4 -----
BRA2 sequence 3.1 TGATTGGGTAAACAAAATGAATGAGTATGAACAGGACATATTAACAACCTCTTCATGCAGC
BRA1 Sequence 2 -----

BRA1 TAAAAATAATTTAGAACAGCGATGATTACTTTTAATACACCTGATAGTATAGCTCAATT
BRA1 Sequence 3 -----
BRA2 sequence 4.1 TAGAAATAATTTAGAACAGGCTATGATTACTTTTAATACACCTGATAGTATAGCTCAATT
BRA1 Sequence 1 -----
BRA2 sequence 5.1 TAGAAATAATTTAGAACAGGCTATGATTACTTTTAATACACCTGATAGTATAGCTCAATT
BRA2 sequence 2.1 TAGAAATAATTTAGAACAGGCTATGATTACTTTTAATACACCTGATAGTATAGCTCAATT
BRA2 sequence 2.2 -----
BRA2 sequence 4.2 -----
BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 TAGAAATAATTTAGAACAGGCTATGATTACTTTTAATACACCTGATAGTATAGCTCAATT
BRA1 Sequence 4 -----
BRA2 sequence 3.1 TAGAAATAATTTAGAACAGGCTATGATTACTTTTAATACACCTGATAGTATAGCTCAATT
BRA1 Sequence 2 -----

BRA1 TGGAAGAAATATTTGGAGTCATATAGCAAACCTGGATTCCGGGACTAGGAGCTTCTATAAT
BRA1 Sequence 3 -----
BRA2 sequence 4.1 CGGAAAAAATATTTGGAGTCATATAGCAAACCTGGATTCCAGGACTGGGAGCTTCAATAAT
BRA1 Sequence 1 -----
BRA2 sequence 5.1 TGGAAAAAATATTTGGAGTCATATAGCAAACCTGGATTCCAGGACTGGGAGCTTCAATAAT
BRA1 Sequence 2 -----
BRA2 sequence 2.1 TGGAAAAAATATTTGGAGTCATATAGCAAACCTGGATTCCAGGACTGGGAGCTTCAATAAT
BRA2 sequence 2.2 -----
BRA2 sequence 4.2 -----
BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 YGGAAAAAATATTTGGAGTCATATAGCAAACCTGGATTCCAGGACTGGGAGCTTCAATAAT
BRA1 Sequence 4 -----
BRA2 sequence 3.1 YGGAAAAAATATTTGGAGTCATATAGCAAACCTGGATTCCAGGACTGGGAGCTTCAATAAT
BRA1 Sequence 2 -----

```

```

BRA1          AAAATATATAATAATGTTGTTATTGTGTTATGTGGTACTAACCTCTTCACCTAAGATCCT
BRA1 Sequence 3 -----
BRA2 sequence 4.1 AAAATATATAGTAATGCTATTATTGTGTTATGTGGTACTAACCTCTTCACCTAAGATCCC
BRA1 Sequence 1 -----
BRA2 sequence 5.1 AAAATATATAGTAATGCTATTATTGTGTTATGTGGTACTAACCTCTTCACCTAAGATCCT
BRA1 Sequence 2 -----
BRA2 sequence 2.1 AAAATATATAGTAATGCTATTATTGTGTTATGTGGTACTAACCTCTTCACCTAAGATCCT
BRA2 sequence 2.2 -----
BRA2 sequence 4.2 -----
BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 AAAATATATAGTAATGCTATTATTGTGTTATGTGGTACTAACCTCTTCACCTAAGATCCT
BRA1 Sequence 4 -----
BRA2 sequence 3.1 AAAATATATAGTAATGCTATTATTGTGTTATGTGGTACTAACCTCTTCACCTAAGATCCT
BRA1 Sequence 2 -----

```

```

BRA1          CAGAAACCTCTTGAGAATGATGAGTGGTGCAGGATCCTCCGTCAATCGCTACCTGAAGAG
BRA1 Sequence 3 -----
BRA2 sequence 4.1 CAGAAACCTCTTGACAACGATGAGTGGTGCAGGATCCTCCGTCAATCACTTCCTGAAGAA
BRA1 Sequence 1 -----
BRA2 sequence 5.1 CAGAAACCTCTTGACAACGATGAGTGGTGCAGGATCCTCCGTCAATCACTTCCTGAAGAA
BRA1 Sequence 2 -----
BRA2 sequence 2.1 CAGAAACCTCTTGACAACGATGAGTGGTGCAGGATCCTCCGTCAATCACTTCCTGAAGAA
BRA2 sequence 2.2 -----
BRA2 sequence 4.2 -----
BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 CAGAAACCTCTTGACAACGATGAGTGGTGCAGGATCCTCCGTCAATCACTTCCTGAAGAA
BRA1 Sequence 4 -----
BRA2 sequence 3.1 CAGAAACCTCTTGACAACGATGAGTGGTGCAGGATCCTCCGTCAATCACTTCCTGAAGAA
BRA1 Sequence 2 -----

```

```

BRA1          ACAACACCATCAGAAACATGCATGGAAAGAAGATGCCTGGGACCAGGACCAGTACAGAAT
BRA1 Sequence 3 -----
BRA2 sequence 4.1 GAAACACCATCAGAAACATGCATGGAAAGAAGAATCCTGGGACCAGGACCAGTACAGTGT
BRA1 Sequence 1 -----
BRA2 sequence 5.1 GAAACACCATCAGAAACATGCATGGAAAGAAGAATCCTGGGACCAGGACCAGTACAGTGT
BRA1 Sequence 2 -----
BRA2 sequence 2.1 GAAACACCATCAGAAACATGCATGGAAAGAAGAATCCTGGGACCAGGACCAGTACAGTGT
BRA2 sequence 2.2 -----
BRA2 sequence 4.2 -----
BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 GAAACACCATCAGAAACATGCATGGAAAGAAGAATCCTGGGACCAGGACCAGTACAGTGT
BRA1 Sequence 4 -----
BRA2 sequence 3.1 GAAACACCATCAGAAACATGCATGGAAAGAAGAATCCTGGGACCAGGACCAGTACAGTGT
BRA1 Sequence 2 -----

```

```

BRA1          ACACCTAGCAGACGCGACCGGTGGCTCAGAGGACAAATACAACAAGCCGAAATGCTCCAA
BRA1 Sequence 3 -----
BRA2 sequence 4.1 ACACCTAGCAGGCGTAACCGGTGGATTAGAGGACAAATACAACAAGCTGAAATACTCCAA
BRA1 Sequence 1 -----
BRA2 sequence 5.1 ACACCTAGCAGGCGTAACCGGTGGATTAGAGGACAAATACAACAAGCTGAAATACTCCAA
BRA1 Sequence 2 -----
BRA2 sequence 2.1 ACACCTAGCAGGCGTAACCGGTGGATTAGAGGACAAATACAACAAGCTGAAATACTCCAA

```

```

BRA2 sequence 2.2 -----
BRA2 sequence 4.2 -----
BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 ACACCTAGCAGGCGTAACCGGTGGATTAGAGGACAAATACAACAAGCTGAAATACTCCAA
BRA1 Sequence 4 -----
BRA2 sequence 3.1 ACACCTAGCAGGCGTAACCGGTGGATTAGAGGACAAATACAACAAGCTGAAATACTCCAA
BRA1 Sequence 2 -----

```

```

BRA1 GAGAAGTTACARTGGCGGATTAAAGGGGTTCAACAGACAGCCAAAGAATTGGAAGAACTC
BRA1 Sequence 3 -----
BRA2 sequence 4.1 GAGAAATTACAATGGCGGATCAAAGGAGTTCAACAAACGGCCCAAGAGCTGGGAGACCTT
BRA1 Sequence 1 -----
BRA2 sequence 5.1 GAGAAATTACAATGGCGGATCAAAGGAGTTCAACAAACGGCCCAAGAGCTGGGAGACCTT
BRA1 Sequence 2 -----
BRA2 sequence 2.1 GAGAAATTACAATGGCGGATCAAAGGAGTTCAACAAACGGCCCAAGAGCTGGGAGACCTT
BRA2 sequence 2.2 -----
BRA2 sequence 4.2 -----
BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 GAGAAATTACAATGGCGGATCAAAGGAGTTCAACAAACGGCCCAAGAGCTGGGAGACCTT
BRA1 Sequence 4 -----
BRA2 sequence 3.1 GAGAAATTACAATGGCGGATCAAAGGAGTTCAACAAACGGCCCAAGAGCTGGGAGACCTT
BRA1 Sequence 2 -----

```

```

BRA1 AACAGAAAGATCAGAAGTGAAGTCTGACTGGACTAGGAGAACACATGGAGACTTCTCATCT
BRA1 Sequence 3 -----
BRA2 sequence 4.1 AATAGACGCATCAGGAAAGAACTCGACTGGACCCGGGCCCAGCATGGGGATTACTCTTCC
BRA1 Sequence 1 -----
BRA2 sequence 5.1 AATAGACGCATCAGGAAAGAACTCGACTGGACCCGGGCCCAGCATGGGGATTACTCTTCC
BRA1 Sequence 2 -----
BRA2 sequence 2.1 AATAGACGCATCAGGAAAGAACTCGACTGGACCCGGGCCCAGCATGGGGATTACTCTTCC
BRA2 sequence 2.2 -----
BRA2 sequence 4.2 -----
BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 AATAGACGCATCAGGAAAGAACTCGACTGGACCCGGGCCCAGCATGGGGATTACTCTTCC
BRA1 Sequence 4 -----
BRA2 sequence 3.1 AATAGACGCATCAGGAAAGAACTCGACTGGACCCGGGCCCAGCATGGGGATTACTCTTCC
BRA1 Sequence 2 -----

```

```

BRA1 TTCCATAGCTATCGCCGAGAACA---GCGTCTCTGGGGACAATGCTCAGAGAACGTCTAC
BRA1 Sequence 3 -----
BRA2 sequence 4.1 TTCGATGGCTATCGGCGAGACCAAGAGCGTCGCTGGGGACAATGCCCAACAAGCATCGGC
BRA1 Sequence 1 -----
BRA2 sequence 5.1 TTCGATGGCTATCGGCGAGACCAAGAGCGTCGCTGGGGACAATGCCCAACAAGCATCGGC
BRA1 Sequence 2 -----
BRA2 sequence 2.1 TTCGATGGCTATCGGCGAGACCAAGAGCGTCGCTGGGGACAATGCCCAACAAGCATCGGC
BRA2 sequence 2.2 -----
BRA2 sequence 4.2 -----
BRA2 sequence 5.2 -----
BRA2 sequence 3.2 -----
BRA2 sequence 6.2 -----
BRA1 Sequence 1 -----
BRA2 sequence 1.1 TTCGATGGCTATCGGCGAGACCAAGAGCGTCGCTGGGGACAATGCCCAACAAGCATCGGC
BRA1 Sequence 4 -----

```

```

BRA2 sequence 3.1   TTCGATGGCTATCGGCGAGACCAAGAGCGTCGCTGGGGACAATGCCCAACAAGCATCGGC
BRA1 Sequence 2     -----

BRA1               AAGTCTGAAGATCCAAAGCGAAGGAGGAAACATCTATGACTGTTGCATCAAAGCCCAAGA
BRA1 Sequence 3     -----
BRA2 sequence 4.1   AAACCTGGAGATTCTGAAGCGAAGGAGGAAACATCTATGACTGTTGCATTAAAGCTCAGGA
BRA1 Sequence 1     -----
BRA2 sequence 5.1   AAACCTGGAGATTCTGAAGCGAAGGAGGAAACATCTATGACTGTTGCATTAAAGCTCAGGA
BRA1 Sequence 2     -----
BRA2 sequence 2.1   AAACCTGGAGATTCTGAAGCGAAGGAGGAAACATCTATGACTGTTGCATTAAAGCTCAGGA
BRA2 sequence 2.2   -----
BRA2 sequence 4.2   -----
BRA2 sequence 5.2   -----
BRA2 sequence 3.2   -----
BRA2 sequence 6.2   -----
BRA1 Sequence 1     -----
BRA2 sequence 1.1   AAACCTGGAGATTCTGAAGCGAAGGAGGAAACATCTATGACTGTTGCATTAAAGCTCAGGA
BRA1 Sequence 4     -----
BRA2 sequence 3.1   AAACCTGGAGATTCTGAAGCGAAGGAGGAAACATCTATGACTGTTGCATTAAAGCTCAGGA
BRA1 Sequence 2     -----

BRA1               GGGGACTCTTGCTATTCCCTGCTGCGGGTTCCCACTATGGCTATTCTGGGGACTCATAAT
BRA1 Sequence 3     -----
BRA2 sequence 4.1   GGGGACTCTTGCTATTCCCTGCTGTGGGTTTCCACTGTGGTTATTCTGGGGATTGTGTGAT
BRA1 Sequence 1     -----
BRA2 sequence 5.1   GGGGACTCTTGCTATTCCCTGCTGTGGGTTTCCACTGTGGTTATTCTGGGGATTGTGTGAT
BRA1 Sequence 2     -----
BRA2 sequence 2.1   GGGGACTCTTGCTATTCCCTGCTGTGGGTTTCCACTGTGGTTATTCTGGGGATTGTGTGAT
BRA2 sequence 2.2   -----
BRA2 sequence 4.2   -----
BRA2 sequence 5.2   -----
BRA2 sequence 3.2   -----
BRA2 sequence 6.2   -----
BRA1 Sequence 1     -----
BRA2 sequence 1.1   GGGGACTCTTGCTATTCCCTGCTGTGGGTTTCCACTGTGGTTATTCTGGGGATTGTGTGAT
BRA1 Sequence 4     -----
BRA2 sequence 3.1   GGGGACTCTTGCTATTCCCTGCT-----
BRA1 Sequence 2     -----

BRA1               CATAATGGGACGTTTGTTTCGGCTATGGGTTGCGAGGGATTGCTATGATACTGTCCTATTG
BRA1 Sequence 3     -----
BRA2 sequence 4.1   CATAATAGGGCGATTATTAGGCTATGGACTGCGAGGCGTTGCTATGATACTGCTCTATTG
BRA1 Sequence 1     -----
BRA2 sequence 5.1   CATAATAGGGCGATTATTAGGCTATGGACTGCG-----
BRA1 Sequence 2     -----
BRA2 sequence 2.1   CATAATAGGGCGATTATTAGGCTATGGACTGCGAGGC-----
BRA2 sequence 2.2   -----
BRA2 sequence 4.2   -----
BRA2 sequence 5.2   -----
BRA2 sequence 3.2   -----
BRA2 sequence 6.2   -----
BRA1 Sequence 1     -----
BRA2 sequence 1.1   CATAATAGGGCGATTATTAGGCTATGGACTGCGRGGCRTYGCTATGATACTGYTCTATTG
BRA1 Sequence 4     -----
BRA2 sequence 3.1   -----
BRA1 Sequence 2     -----

BRA1               TGAAAAAGGATTGACTATGCTATTTATAATGCTAAA
BRA1 Sequence 3     -----
BRA2 sequence 4.1   TGGGAAAGGATTAACATATGCTGTTTATTATGTAAAA
BRA1 Sequence 1     -----

```

BRA2	sequence	5.1	-----
BRA1	Sequence	2	-----
BRA2	sequence	2.1	-----
BRA2	sequence	2.2	-----
BRA2	sequence	4.2	-----
BRA2	sequence	5.2	-----
BRA2	sequence	3.2	-----
BRA2	sequence	6.2	-----
BRA1	Sequence	1	-----
BRA2	sequence	1.1	TGGGAAAGGATTAAGTATGCTGTTTATTA-----
BRA1	Sequence	4	-----
BRA2	sequence	3.1	-----
BRA1	Sequence	2	-----

Appendix C: Alignment of nucleotide sequences in EIAV LTR using MUSCLE. U3 is in black, R in red, and U5 in blue. Methylated DNA-binding protein site (MDBP) is in the red box, PEA1/AP-1 site in pink, transactivation response element (TAR) is in the green box, initiation site of transcription is in gray, and PU.1 binding sites is in yellow.

```

BRA1      TGTGGGGTTAATAT-AAAATGGCTGTAGAAAAGAAAGTTGCTGATGATTATAAGTA-----AATATGCTGA
BRA2      .....G.....A.....G.....
Liaoning   .....A.....C...G...TCT.A...-AG.....GT.....
KY411969   .....A.....-..G...TCT.A...-AG.....GT.....
KY411970   .....A.....-..G...TCT.A...-AG.....GT.....
KY411971   .....A.....-..G...TCT.A...-AG.....GT.....
KY411972   .....A.....-..G...TCT.A...-AG.....GT.....
KY411973   .....A.....-..G...TCT.A...-AG.....GT.....
KY411974   .....A.....-..G...TCT.A...-AG.....GT.....
KY411975   .....A.....-..G...TCT.A...-AG.....GT.....
KY411976   .....A.....-..G...TCT.A...-AG.....GT.....
KY412097   .....A.....-..G...TCT.A...-AG.....GT.....
KY412098   .....A.....-..G...TCT.A...-AG.....GT.....
DLV9-7     .....A.....-..G...TCT.A...-AG...A...G.AAGTTGCTGATGCTCTC.GT.....
DV10-3     .....A.....-..G...TCT.A...-AG...A...G.....GT.....
IRE_F2     .....A...-ATAA.G-----AG.AC..G...AA.GA-----CT.....
IRE_F3     .....A...-ATAA.G-----AG.AC..G...AA.GA-----CT.....
IRE_F4     .....A...-ATAA.G-----AG.AC..G...AA.GA-----CT.....
IRE_H3     .....A...-ATAA.G-----AG.AC..G...AA.GA-----CT.....
Wyoming    .....TT...G.GGGGTTT.A...-A.....AGTAAAAAG-----A..GT.....
Miyazaki   ...A.A...-A...-TG.AT..G..TA-----AC.....

```

```

BRA1      TGCTCT-CATAACCGCGTGTAACTAGAAAG-CTA-GCTCATGTTGCTAGGCAACTAAACTGTAATAACCT-GTTAGTTCC
BRA2      .....-.....-A.....A...A.....A.....-.....T.....
Liaoning   .....-.....TTA.....C.A...A.....C.....G.....-.....T.....
KY411969   .....-.....TTA.....C.A...A.....C.....-.....T.....
KY411970   .....-.....TTA.....C.A...A.....C.....-.....T.....
KY411971   .....-.....TTA.....C.A...A.....C.....-.....T.....
KY411972   .....-.....TTA.....C.A...A.....C.....-.....T.....
KY411973   .....-.....TTA.....C.A...A.....C.....-.....T.....
KY411974   .....-.....TTA.....C.A...A.....C.....-.....T.....
KY411975   .....-.....TTA.....C.A...A.....C.....-.....T.....
KY411976   .....-.....TTA.....C.A...A.....C.....-.....T.....
KY412097   .....-.....TTA.....C.A...A.....C.....G.....-.....T.....
KY412098   .....-.....TTA.....C.A...A.....C.....C.....CG..C.....-.....T.....
DLV9-7     .....-.....TTA.....-A.....C.....G.....G.....-.....T.....
DV10-3     .....-.....TTA.....C.AG..A.....C.....G.....-.....T.....
IRE_F2     .....-T.....T.A.....A...G.....G.....A...G.....-.....A.....
IRE_F3     .....-T.....T.A.....A...G.....G.....A.....-.....A.....
IRE_F4     .....-T.....T.A.....A...G.....G.....A...G.....-.....A.....
IRE_H3     .....-T.....T.A.....A...G.....G.....A...G.....-.....A.....
Wyoming    .....-.....TT..A.....C.A.G.A.....C.....GCA..T..GA..
Miyazaki   .....A.....TT..A.....C.T.TAA..G..AGGC.....A.....G.....AAT.....-A.....

```

```

BRA1      TCATTATTGT-----TCCTGTTTTTACACATATAGTACATGTAT
BRA2      .....-.....-.....G.....G.....
Liaoning   .....A...-TCCGCTTTTGTGACGCGTTAAGT...G.....T.....
KY411969   .....A...-TCCGCTTTTGTGACGCGTTAAGT...G.....G.T.....
KY411970   .....A...-TCCGCTTTTGTGACGCGTTAAGT...G...C.....G.T.....
KY411971   .....A...-TCCGCTTTTGTGACGCGTTAAGT...G.....G.T.....
KY411972   .....A...-TCCGCTTTTGTGACGCGTTAAGT...G.....G.T.....
KY411973   .....A...-TCCGCTTTTGTGACGCGTTAAGT...G.....G.T.....
KY411974   .....A...-TCCGCTTTTGTGACGCGTTAAGT...G.....G.T.....
KY411975   .....A...-TCCGCTTTTGTGACGCGTTAAGT...G.....G.T.....
KY411976   .....A...-TCCGCTTTTGTGACGCGTTAAGT...G.....G.T.....
KY412097   .....A...-TCCGCTTTTGTGACGCGTTAAGT...G.....T.....
KY412098   .....A...-TCCGCTTTTGTGACGCGTTAAGT...G.....T.....
DLV9-7     .....A...TCCGCTTTTGTATAGTTCCGCTTTTGTGACGCGTTAAGT...G.....G.T.....
DV10-3     .....A...-TCCGCTTTTGTGACGCGTTAAGT...G.....G.T.....
IRE_F2     .....A...-TCCGCTTTTGTGACGCGTTAAGT...CAAA...A...C..T.....
IRE_F3     .....A...-TCCGCTTTTGTGACGCGTTAAGT...CAAA...A...C..T.....
IRE_F4     .....A...-TCCGCTTTTGTGACGCGTTAAGT...CAAA...A...C..T.....
IRE_H3     .....A...-TCCGCTTTTGTGACGCGTTAAGT...CAAA...A...C..T.....
Wyoming    G.G---A...-TCCCCATTGGTGACGCGTTAACT...G.....G.T.....
Miyazaki   CT...TAA...-TCCGCTTTTGTGACGCGTTAAGT...A.....A...G.T.A...

```

```

BRA1      TCTGATAACAAGACACTCAGATTCTGCGGTCTGAGTCCCTTCTCTGCTGGGCCTAATA--AGCCTTGA-AATAA--ATAT
BRA2      .T..T.....C.--.....A.G.....
Liaoning  ....C.TTTG.....T-.CTCT.....GT.....
KY411969  ....C.TTTG.....T-G.CTCT.....GT.....
KY411970  ....C.TTTG.....T-G.CTCT.....GT.....
KY411971  ....C.TTTG.....T-G.CTCT.....GT.....
KY411972  ....C.TTTG.....T-G.CTCT.....GT.....
KY411973  ....C.TTTG.....T-G.CCCT.....GT.....
KY411974  ....C.TTTG.....T-G.CTCT.....GT.....
KY411975  ....C.TTTG.....T-G.CTCT.....GT.....
KY411976  ....C.TTTG.....T-G.CTCT.....GT.....
KY412097  ....C.TTTG.....T-.CTCT.....GT.....
KY412098  ....C.TTTG.....T-.CTCT.....GT.....
DLV9-7    ....C.TTTG..T.G.....TA..CTCT.....GT.....
DV10-3    ....C.TTTG.....TA..CTCT.....GT.....
IRE_F2    ....C.CTGG.....TA..A---.....CTT.....
IRE_F3    ....C.CTGG.....TA..A---.....CTT.....
IRE_F4    ....C.CTGG.....TA..A---.....CTT.....
IRE_H3    ....C.CTGG.....TA..A---.....CTT.....
Wyoming   ....C..TTG..G-.....TG..A.--G.....TGT.....
Miyazaki  ....TC.TTTG..G.....TA.....C.....AT.....

```

```

BRA1      AATTCTCTA-CTCAGTCCCTTGCTCTCAACTTGT--TGGTTTTAA-GATCCTACA
BRA2      .....GA.....A.....-T..A.....-.....
Liaoning  .....G-.....GT...T.GTC...CT..T..C..-..G..TA...
KY411969  .....G-T.....GT...T.GT...CT..T..C..-..TA...
KY411970  .....G-T.....GT...T.GT...CT..T..C..-..TA...
KY411971  .....G-T.....GT...T.GT...CT..T..C..-..TA...
KY411972  .....G-T.....GT...T.GT...CT..T..C..-..TA...
KY411973  .....G-T.....GT...T.GT...CT..T..C..-..TA...
KY411974  .....G-T.....GT...T.GT...CT..T..C..-..TA...
KY411975  .....G-T.....GT...T.GT...CT..T..C..-..TA...
KY411976  .....G-T.....GT...T.GT...CT..T..C..-..TA...
KY412097  .....CG-.....GT...T.GTC...CT..T..C..-..TA...
KY412098  .....G-.....GT...T.GTC...CT..T..C..-..TA...
DLV9-7    .....G-..A.....GT...T.GT...CT..T..C..-..TA...
DV10-3    .....G-..A.....GT...T.GT...CT..T..C..-..TA...
IRE_F2    .....-.....T...GTGT.....A.CTCTTGG.....-.....
IRE_F3    .....-.....T...GTGT.....A.CTCTTGG.....-.....
IRE_F4    .....-.....T...GTGT.....A.CTCTTGG.....-.....
IRE_H3    .....-.....T...GTGT.....A.CTCTTGG.....-.....
Wyoming   .....-.....GTCTCT.GT...C---TG..CG.-.....
Miyazaki  .....G-T.....GTGT.....TT..C..AC.GCA.C..A...

```

Appendix D: Alignment of the gag polyprotein. The proteins p15, p26, p11 and p9 are separated by blue arrows. Amino acid variations between BRA1 and BRA2 are in yellow. A leucine-rich-type nuclear export signal is in the green box (p15), major homology region is in the red box (p26), two zinc binding domains are in the blue boxes (p11), and YPDL late domain is in the orange box (p9).

```

BRA1      MGDPVITWSKALKKLEKVTVSGSQKLNSGNCNWALSLVDLYHDTNYLKEKDWQLRNVIPLLEDVTQTLSGQ
BRA2      ...L.....TT.....F.....FT.....D.....S.....L
DLV9-7_HM141922 ...SL.....Q.....T.....N.....F.....FC.....D.....V.....S.....
DV10-3_HM141910 ...SL.....Q.....T.....N.....F.....FC.....D.....S.....
IRE_F2_JX480631 ...KAS..L.....L.....FS.....DIL.....S.....T..
IRE_F3_JX480632 ...KAS..L.....L.....FS.....DIL.....S.....T..
IRE_F4_JX480633 ...KAS..L.....L.....FS.....DIL.....S.....T..
IRE_H3_JX480634 ...KAS..L.....L.....FS.....DIL.....S.....T..
ITA_DE_KM247554 ...KAS..L.....L.....FS.....DIL.....S.....T..
ITA_SA_KM247555 ...KAS..L.....L.....FS.....DIL.....S.....T..
Liaoning_AF327877 -----
Miyazaki_JX003263 ...SLN.....Q.....QP.....F.....FM.....KD.....AE....P
Wyoming_AF033820 ...L.....Q.....TT.....F.....FV.....D.....

```

```

BRA1      EREAFEKTWWAISAVKMGMQINNVDGKASFQLLKAKYDKRD--TGKKQPEPPEEFIPIMIDGAGNRNFRP
BRA2      ...L.....AG.....V.....E.KA--R.....Y.....
DLV9-7_HM141922 ...A.....L.....T.N.A.TT.SI....FERKTVN.T...S..M..Y.....
DV10-3_HM141910 ...A.....L.....T.N.A.TT.TI....FERKTAN.T...S..E..Y.....
IRE_F2_JX480631 ...K.....V...K.L.GH.G..R.TY.....ERAP--SK.QH...Q..Y.....
IRE_F3_JX480632 ...K.....V...K.L.GH.G..R.TY.....ERAP--SK.QH...Q..Y.....
IRE_F4_JX480633 ...K.....V...K.L.GH.G..R.TY.....ERAP--SK.QH...Q..Y.....
IRE_H3_JX480634 ...K...R...V...K.L.GH.G..R.TY.....ERAP--SK.QH...TQ..Y.....
ITA_DE_KM247554 ...K.....V...K.L.GH.G..R.TY.....ERAP--SK.QH...Q..Y.....
ITA_SA_KM247555 ...K.....V...K.L.GH.G..R.TY.....ERAP--SK.QH...Q..Y.....
Liaoning_AF327877 -----L...T.N.A..T.TI....FERKTAN.T...S..M..Y.....
Miyazaki_JX003263 ...K...R...T.LM...H...T.....G.GA--...TN.Q.AY..IT..G...T..
Wyoming_AF033820 ...R.....L.....V.....R...E.KT--AN...S..S..Y.....

```

```

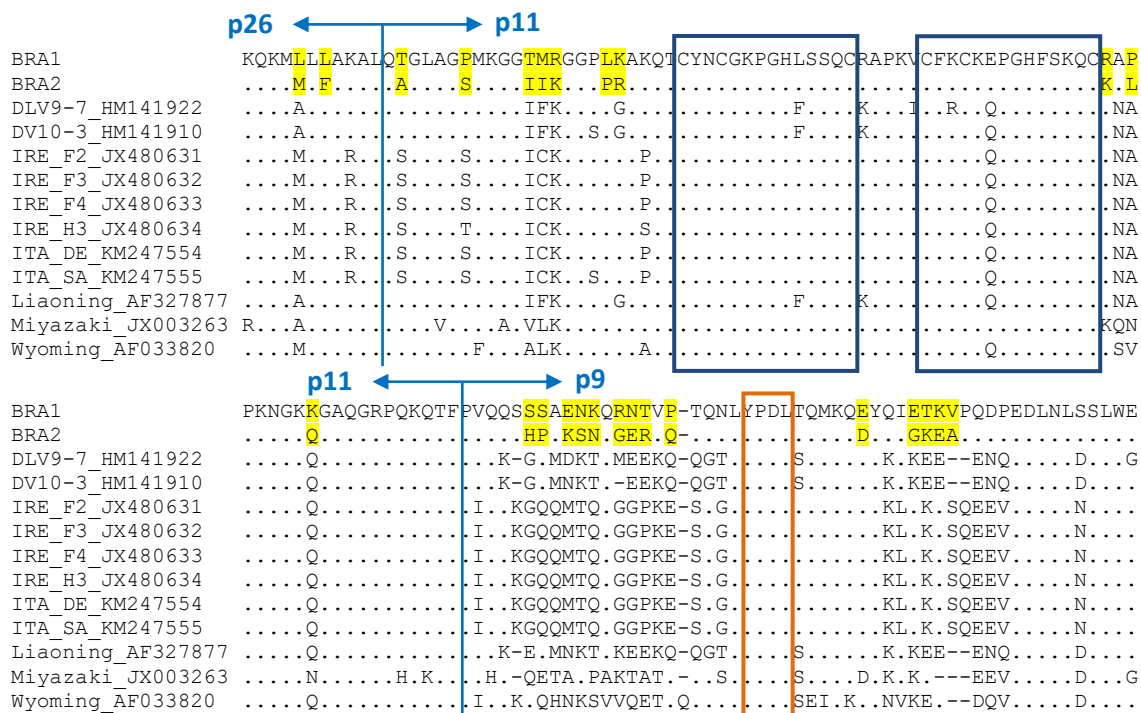
BRA1      LTPRGYTTWVNTIQQNNLLNEASVNLFGILSVDCTSEEMNAFLDVVPGAGQKQVLLDLLDKMAEDWDNR
BRA2      .....I.....
DLV9-7_HM141922 .....N...V..E..R..
DV10-3_HM141910 .....N...I..E..R..
IRE_F2_JX480631 .....I.Q...A
IRE_F3_JX480632 .....E.I.Q...A
IRE_F4_JX480633 .....I.Q...A
IRE_H3_JX480634 .....I.Q...A
ITA_DE_KM247554 .....E.I.Q...A
ITA_SA_KM247555 .....E.I.Q...A
Liaoning_AF327877 .....N...I..E..R..
Miyazaki_JX003263 .....S...T.....I.....N.E.I..E..R..
Wyoming_AF033820 .....T.G.....Q.....I...AI..I.D....

```

```

BRA1      HPLPNPPLVAPAQGPIPMRTARFIRGLGVPRERQMEPAFDQFRQTYRQWIIIEAMTEGIKIMIGKPKAQNIR
BRA2      .....M.....
DLV9-7_HM141922 ...M.....P.....M.....
DV10-3_HM141910 ...M.....P.....
IRE_F2_JX480631 ...A.M...P.....
IRE_F3_JX480632 ...A.M...P.....
IRE_F4_JX480633 ...A.M...P.....
IRE_H3_JX480634 ...A.M...P.....
ITA_DE_KM247554 ...A.M...P.....
ITA_SA_KM247555 ...A.M...P.....
Liaoning_AF327877 ...M.....P.....
Miyazaki_JX003263 N.I...M...P.....
Wyoming_AF033820 ...A....P.....S...V.....

```



Appendix E: Alignment of *pol* polyprotein. In PR, the catalytic motif DTG is in the red box, substrate-binding sites are in cyan and active sites flap are in the box orange. In RT, DIGD motif is in the purple box and DD motif is in the gray box. In RNase H, invariant residues are in cyan and the glycine-rich motif is in the dark green box. In DU, the conserved motifs MI-MV are in the dark blue boxes, with the conserved residue D72 in MIII in cyan. In IN, the zinc finger-binding motif (H12, H16, C40, C43) are in green, hydrophobic core residues (I5, A8, L22, A33 and I36) are in pink, and central catalytic core residues (D64, D116 and E152) are in blue.

	Substrate binding																		
	DTG catalytic motif										residues			Active site flap					
BRA1	VGVTYNLQKR	RPT	SIVL	INDTPL	SVLL	DTG	ADTS	SVLT	IAHYNK	LKYR	GKKY	QGT	SIVGV	VGCNV	ET	FSTP	VTIK	KKGK	QFRT
BRA2	E.....	A.....							R.....										IK.
DLV9-7	E.....	T.....			N.....				R.....		R.....						V.....		IK.
DV10-3	E.....	T.....			N.....				R.....		R.....						V.....		IK.
IRE_F2	E.....	TV.....			N.....			V.....			R.....								IK.
IRE_F3	E.....	TV.....			N.....			V.....			R.....								IK.
IRE_F4	E.....	TV.....			N.....			V.....			R.....								IK.
IRE_H3	E.....	TV.....			N.....			V.....			R.....								IK.
ITA_DE	E.....	TV.....			N.....			V.....			R.....								IK.
ITA_SA	E.....	TV.....			N.....			V.....		K.R.....									IK.
Liaoning	E.....	T.....			N.....				R.....		R.....						V.....		IK.
Miyazaki	K.DR.....	TV.I.....			N.....				T.S.....		R.....			L.....					IK.
Wyoming	E.....	T.....			N.....				T.....		R.....				I.....				RHIK.

PR

RT

BRA1	RMI	VADIPVT	ILGRDILQELGAQLVMA	QLSKEITPKKIKIKEGMS	GP	KIPQWPLTKEKLVGAKEIVQRL	LLSEGKISEASS
BRA2	..I.....			R.....		I.....	
DLV9-7	K.....		I.....	RE..L.T.TV.....	..V.....	L.....	KK..D.....D
DV10-3			I.....	RE..L.T.TV.....	..V.....	L.....	KK..D.D.....D
IRE_F2			V.....	V.R..L.T.TT.....		L.....	L.....ED
IRE_F3			V.....	V.R..L.T.TT.....		L.....	L.....ED
IRE_F4			V.....	V.R..L.T.TT.....		L.....	L.....ED
IRE_H3			V.....	V.R..L.T.TT.....		L.....	L.....ED
ITA_DE			V.....	V.R..L.T.TT.....		L.....	L.....ED
ITA_SA			V.....	V.R..L.T.TT.....		L.....	L.....ED
Liaoning			I.....	RE..L.T.TV.....	..V.....	L.....	KK..D.....D
Miyazaki			I.....	T.S.RR..L.P..T.....		T.....	I.....D
Wyoming			D..K..L.....	KFR..EL..TM.....		E.....	D

DIGD motif

BRA1	ENPYNSPIFVIRKKSGKWRL	LIQDLRALNTIVKVGTEISRGLPHPGGLIKSNHMTVLD	DIGL	AYFTIPLDPEFRPYTAFTVP
BRA2	L.....V.....Q.....	ND.....		
DLV9-7	D.....K.....L.....E..KV.Q.....	C.....		K..Q.....
DV10-3	D.....K.....L.....E..KV.Q.....	C.....		K..Q.....
IRE_F2	N.....K.....L.....E..KM.EL.....	CKQ.....		
IRE_F3	N.....K.....L.....E..KM.EL.....	CKQ.....		
IRE_F4	N.....K.....L.....E..KM.EL.....	CKQ.....		
IRE_H3	N.....K.....L.....E..KM.EL.....	CKQ.....		
ITA_DE	N.....K.....L.....E..KM.EL.....	CKQ.....		
ITA_SA	N.....K.....L.....E..KM.EL.....	CKQ.....		
Liaoning	D.....K.....L.....E..KV.Q.....	C.....		K..Q.....
Miyazaki	D..F.....K.R.....L.....E..KA.Q.....	KN..M.....		L..R.....
Wyoming	N.....K.R.....L.....E..KT.Q.....	CK.....		I.....

DD motif

BRA1	SLNHQEPDKRYVWNCLPQG	FVLSPYIYQKTLQEI	LQAFRERYPEVQLYQYM	DL	FIGSNGSKAQHKELIKELRTILLEKG
BRA2				V.....R.....N.....	
DLV9-7	..I.....	D.....	H.D.....	E..R.....V.....A.....	
DV10-3	..I.....	D.....	H.D.....	E..R.....VE.....A.....	
IRE_F2	..I.....I.....		HS.....	V..K.....V.....A.....	
IRE_F3	..I.....I.....		HS.....	V..K.....V.....A.....	
IRE_F4	..I.....I.....		HS.....	N.V..K.....V.....A.....	
IRE_H3	..I.....I.....		HS.....	V..K.....V.....A.....	
ITA_DE	..I.....I.....		HS.....	V..K.....V.....A.....	
ITA_SA	..I.....I.....		HS.....	V..K.....V.....A.....	
Liaoning	..I.....I.....	N.....	H.D.....	R.....VE.....A.....	
Miyazaki	..I.....	I.....	P..D.H.....	E.RK..N..V.....M.....	
Wyoming	..I.....K.....		P.....	M.....K.....I.....A.....	

RT ←

BRA1	FETPDDKVQEEAPYNWLG	YQIKPSSSWKVQKMQLEMVQ	EPTLNDVQKLMGNITWMSSGVPGLTVKQIAATT	KGCLDLNQKV
BRA2				
DLV9-7	...E..L.....	...S.GN.....	...L.K.....	
DV10-3	...E..L.....	...S.GN.....	...L.K.....	
IRE_F2	L..P.....	...C.DN.....	...L.K.....	
IRE_F3	L..P.....	...C.DN.....	...L.K.....	
IRE_F4	L..P.....	...C.DN.....	...L.K.....	
IRE_H3	L..P.....	...C.DN.....	...L.K.....	
ITA_DE	L..P.....	...C.DN.....	...L.K.....	
ITA_SA	L..P.....	...C.DN.....	...L.K.....	
Liaoning	...E..L.....	...S.GN.....	...L.K.....	
Miyazaki	...EE..L.D...G.....	...F.NN..MPTV...	...LAK.....	...I...I...H...E...
Wyoming	L..VP..S.....	...C.EN.....	...D..KN.....	...I...H...E...

BRA1	TWTEEAQKELEENNEKIRNAQGLQYYDPKEE	IICEIELTKNYEAIYMIKQPQGILWAGKKIMKANKGWSAAKNLMLLLQH
BRA2	V.....R.....K.....	L.....H.....R.....
DLV9-7	V.....K..QE.....	...N.E..V...I...T.I...S.....R.....
DV10-3	V.....K..QE.....	...N.E..V...I...T.I...S.....R.....
IRE_F2	I..K...I.....K.....	...H..N.E..T...I...IV..S.K.....R.....T.....
IRE_F3	I..K...I.....K.....	...H..N.E..T...I...IVR..S.K.....R.....VV.....
IRE_F4	I..K...I.....K.....	...H..N.E..T...I...IVR..S.K.....R.....
IRE_H3	I..K...I.....K.....	...H..N.E..T...I...IVR..S.K.....R.....T.....
ITA_DE	I..K...I.....K.....	...H..N.E..T...I...IVR..S.K.....R.....
ITA_SA	I..K...I.....K.....	...H..N.E..T...I...IVR..S.K.....R.....
Liaoning	V.....K..QE.....	...N.E..V...I...T.I...S.....R.....
Miyazaki	N..A...Q...Q...VQK.....	...N.D..V...N...C..T.IV..S.....R..TV.....
Wyoming	I.....K.....	...N.E..ML..V.I...T.V...S.....TV.....

→ **RN**

BRA1	VATESITRVGVC	PKFKVPFTKEQVMWEMQKGWYYSWLPEIIYSNQVVHDSWRMKLVEQPTAGIT	TIYT	DGGKKN	GE	GIAAY
BRA2						
DLV9-7	V..I..T.....	...K...E.....DM...H...D..L.....S.....				
DV10-3	V..I..T.....	...K...E.....DM...H...D..KL.....S.....				
IRE_F2		I.....A.....S...F..D..KL.....S.....				
IRE_F3		I.....A.....S...F..D..KL.....S.....				
IRE_F4		I.....A.....S...F..D..KL.....S.....				
IRE_H3		I.....A.....S...F..D..KL.....S.....				
ITA_DE		I.....A.....S...F..D..KL.....S.....				
ITA_SA		I.....A.....S...F..D..KL.....S.....				
Liaoning	V..I..T.....	...K...E.....DMV..H...D..KL.....S.....				
Miyazaki	I..I..T.....	...S...K.....S...E..KL...E..S.....				
Wyoming	K...T.....	V..TH...D.....E..S.....				

Glycine-rich motif

BRA1	VTSDGKTKQRQLGLTTHQAA	ERIAIQMALEDIRD	KQVNI	VTDSYYCWNK	ITEGLGLEGP	DSPPWWPIICNIQDKDAVYFAW
BRA2	PV...T...V.....	K.....	F.....			
DLV9-7	...N....KR..PV...T.....	TEETL.....				...Q..RA.EM...
DV10-3	...N....KR..PV...T.....	TEETL.....				...Q..RA.EM...
IRE_F2	I..K..K..KL..PV...T.....	SEE..I.....				...Q..KT.E.I...
IRE_F3	I..K..K..KL..PV...T.....	SEE..I.....				...Q..KT.E.I...
IRE_F4	I..K..K..KL..PV...T.....	SEE..I.....				...Q..KT.E.I...
IRE_H3	I..K..K..KL..PV...T.....	SEE..I.....				...Q..KT.E.I...
ITA_DE	I..K..K..KL..PV...I.....	SEE..I.....				...Q..KT.E.I...
ITA_SA	I..K..K..KL..PV...I.....	SEE..I.....				...Q..KT.E.I...
Liaoning	...N....KR..PV...T.....	TEETL.....				...Q..RA.EM...
Miyazaki	...T....K...PV...R...M.....	YN...L.....				...Q..YN.ETI...
Wyoming	...N.R....KR..PV...V...M.....	T.....				...Q..RE.EI...

RT ← → **DU** **MI**

BRA1	VPGHKG	IYGNQLAD	DATKIVQEI	MIAYKGTQIKEK	RDEDA	AGFDLCIPYDMTIK	VLETKIISTDVRIQVPPNSFGWVTGKS
BRA2							
DLV9-7	E.....	TE..L..Q.....				II.P.S...V.Y...K...QKC.....	
DV10-3	E.....	TE..L..Q.....				I.P.S...V.P...K...HKC.....	
IRE_F2	E.....	MSE..L..Q...R.....				H..ILHTG...P...K...QQ.....	
IRE_F3	E.....	MSE..L..Q...R.....				H..ILHTG...P...K...QQ.....	
IRE_F4	E.....	MSE..L..Q...R.....				H..ILHTG...P...K...QQ.....	
IRE_H3	E.....	MSE..L..Q...R.....				H..ILHTG...P...K...QQ.....	
ITA_DE	E.....	MSE..L..Q...R.....				H..ILHTG...P...K...QQ.....	
ITA_SA	E.....	MSE..L..Q...R.....				H..ILHTG...P...K...QQ.....	
Liaoning	E.....	TED..L..Q.....				II.P.S...V.P...K...HKC.....	
Miyazaki	E.....	TE..L..Q...RN..K.....				S...I.LH.S...P...K...QC.....	
Wyoming	E.A.....	KE..L..Q.....				V...IM.P.SD...P...K.....	

	MII	MIII	MIV	MV	DU	IN
BRA1	SMKQGLLINGSIIDEGYTGEIQVICTNIGKSNIKLLESGKFAQLIILQHQSNIHVQPWDENKKSQRGEEGFGSTQMEFWVE					
BRA2						
DLV9-7						
DV10-3						
IRE_F2						
IRE_F3						
IRE_F4						
IRE_H3						
ITA_DE						
ITA_SA						
Liaoning						
Miyazaki						
Wyoming						

Hydrophobic core residues

Zinc finger binding motif

BRA1	NIQEAQDEENWITSPKILAKRYGLPIAVAKOITQEPHETKQSGGPAGCVMRSPNHWQACTHIENKVILTFVESNSGY
BRA2	
DLV9-7	
DV10-3	
IRE_F2	
IRE_F3	
IRE_F4	
IRE_H3	
ITA_DE	
ITA_SA	
Liaoning	
Miyazaki	
Wyoming	

Central catalytic core residues

BRA1	IHATLLSKENALCTSLAILEWARLFSPRSLHTNGTNFIADTVQNLLRFLKITHHTGIPYHPESQGIIERANRTLKEKIQ
BRA2	
DLV9-7	
DV10-3	
IRE_F2	
IRE_F3	
IRE_F4	
IRE_H3	
ITA_DE	
ITA_SA	
Liaoning	
Miyazaki	
Wyoming	

BRA1	SHRDNTQTLEAALQLALITCNKGRESMGQTTPWEVFITNQAQVIHDELLLQQAQSSKKKFCFYKIPGESDWKGPTKVLWKG
BRA2	
DLV9-7	
DV10-3	
IRE_F2	
IRE_F3	
IRE_F4	
IRE_H3	
ITA_DE	
ITA_SA	
Liaoning	
Miyazaki	
Wyoming	

BRA1	DGAVVVNDEGKGIIAVPLTRTKLLIKPH
BRA2	
DLV9-7	
DV10-3	
IRE_F2	
IRE_F3	
IRE_F4	
IRE_H3	
ITA_DE	
ITA_SA	
Liaoning	
Miyazaki	
Wyoming	

Appendix F: Alignment of amino acid sequences in EIAV env gene. The surface protein and transmembrane protein were aligned using MUSCLE. Hypervariable regions are marked with colors: yellow - V1, green - V2, cyan - V3, blue - V4, orange - V5, purple - V6, dark green - V7, and red - V8. While cysteine residues are with red letters.

```

BRA1      MVSIIFYGGI PGGISTPIIQ S--SEEASHT KGD-TIFKPY CSDEKSNRNK TMAEERD---
BRA2      ..... ..V.A..V. --P...TC. ...-..... YNDN...KD. K...G.---
DLV9-7_HM141922 ..... ..E....T. Q--T.STDTQ ...HVVYQ... YNDSHKKE- ...A.---
DV10-3_HM141910 ..... ..V....T. Q--T.STDTQ ...HVVYQ... YNDSHKAE- ...A.---
IRE_F2_JX480631 ..... ..VQ...V. QSEP.SKGSE .R.YE..Q... Y.SSYTK- ...G.---
IRE_F3_JX480632 ..... ..VQ...V. QSEP.SKGSE .R.YE..Q... Y.SSYTK- ...G.---
IRE_F4_JX480633 ..... ..VQ...V. QSEP.SKGSE .R.YE..Q... Y.SSYTK- ...G.---
IRE_H3_JX480634 ..... ..VQ...V. QSEP.SKGSE .R.YE..Q... Y.SSYTK- ...G.---
ITA_DE_KM247554 ..... ..VQ...V. QSEP.SKGSE .R.YE..Q... Y.SSYKTE- ...G.---
ITA_SA_KM247555 ..... ..VQ...V. QSEP.SKGSE .R.YK..Q... Y.SSYKTE- ...G.---
Liaoning_AF327877 ..... ..T...T. Q--TKSTDQ ...HVVYQ... YNDSHKAE- ...A.---
Miyazaki_JX003263 ..... ..SS E--..Q.DTD .RGPMV.Q... YNGSNK.R- ...GKQIP
Wyoming_AF033820 ..... ..T...T. Q--..KSKCE EN--..M.Q... YNND.KNS- ...SKE---

```

```

BRA1      SRERREAYLK EDAENAKKR NNDWWKIGMF MLC LMGATGG MLWWYEDVAH PQYIGLITVG
BRA2      P.D...MN.. ..P..K.R.. ..... ..T....I.....G....R.....
DLV9-7_HM141922 T.YQE.MNQ. .EK.D--.. R.N..... L...L.T... F....RQQ. SY....V.I.
DV10-3_HM141910 T.YQE.MNR. .K.D--.. R.N..... L...L.T... F....GQH. SH....V.I.
IRE_F2_JX480631 ..YQE.MIP. .ESKGKEE.G R..... ..L..A.. I....GTPD VH....VA..
IRE_F3_JX480632 ..YQE.MIP. .ESKGKEE.G R..... ..L..A.. I....GTPD VH....VA..
IRE_F4_JX480633 ..YQE.MIP. .ESKGKEE.G R..... ..L..A.. I....GAPD VH....VA..
IRE_H3_JX480634 ..YQE.MIP. .ESKGKEE.G RK..... ..L..A.. I....GTPD VH....VA..
ITA_DE_KM247554 ..YQE.MIP. .ESKGKEE.G R..... ..L..A.. I....GAPD VH....VA..
ITA_SA_KM247555 ..YQE.MIP. .ESKGKEE.G R..... ..L..A.. I....GAPD VH....VA..
Liaoning_AF327877 T.YQE.MNR. .EK.D--.. R.N..... L...L.T... F....GQK. SH....V.I.
Miyazaki_JX003263 PE..E.LFQG .ELKK---E. R..... L...L.T... F....GQE. SH...PVA..
Wyoming_AF033820 A.DQ-.MN.. .ESK.E--.. R..... L...A.T... I....GLPQ QH....VAI.

```

```

BRA1      GKLESGMTS AIECWGKFPG CRPFDNYFSY MTNRRIHVQN NT-VSLESY HREITYIYKT
BRA2      ....E..... .....T... ..T....E...T.QET. D.-AT.... N??-----
DLV9-7_HM141922 .R.N..... .....S... ..T....E...TVSRD. --AT..DA. Q..V.N..R.
DV10-3_HM141910 .R.N..... .....S... ..T....E...TVSRD. --AT..DA. Q...N..R.
IRE_F2_JX480631 .R.N.....N .....T... ..T....K...S..AD. --AT..NA. ....
IRE_F3_JX480632 .R.N.....N .....T... ..T....K...S..AD. --AT..NA. ....
IRE_F4_JX480633 .R.N.....N .....T... ..T....K...S..AD. --AT..NA. ..V.....
IRE_H3_JX480634 .R.N.....N .....T... ..T...R.K...S..AD. --AT..NA. ..V.....
ITA_DE_KM247554 .R.N.....N .....T... ..T....K...S..AD. --AT..NA. ..V.....
ITA_SA_KM247555 .R.N.....N .....T... ..T....K...S..AD. --AT..NA. ..V.....
Liaoning_AF327877 .R.N..... .....S... ..T....E...T.SRD. --AT..DA. Q..V.N..R.
Miyazaki_JX003263 .G.N..... .....SS... ..N....E...K.T.NID. --AAT..QA. ..K.G.L..
Wyoming_AF033820 .R.N...QSN .....S... ..Q....E...SM.MD. --AT..A. ....F...S

```

```

BRA1      MCTDSDHCQE YECRNVLQTLG NGLKVNISYN TTLDTP---Y WNFKWLVCNQ TENQKAILIP
BRA2      -.....? ...ER.R.N. -NFSDR..VE DNNNST---. .D.T..L... ..K.T..V.
DLV9-7_HM141922 S.V.....K .K.KQ...E- -KNSN..IM. NCSNNSCEEF .G.S..E... ..SAI...V.
DV10-3_HM141910 S.V.....K .K.KQ...K- -ENSS..IM. NCSSNSCEEF RG.S..E... ..AIT..V.
IRE_F2_JX480631 S.V.....V .V.QV.DM.- -DKNNTLTIS I.GNGT--D. .G.Q..E... ..A.T.VV.
IRE_F3_JX480632 S.V.....V .V.QV.DM.- -DKNNTLTIS I.GNGT--D. .G.Q..E... ..A.T.VV.
IRE_F4_JX480633 S.V.....V .V.QISM.- -NKGSTLTIS I.GGT--D. .G.Q..E... ..A.TMVV.
IRE_H3_JX480634 S.V.....V .V.QISM.- -DNGN..II- -.GNGT--D. .G.Q..E... ..A.T.VV.
ITA_DE_KM247554 S.V.....V .V.QA.NM.- -DGGNTLTIS I.GNRT--D. .G.R..E... ..A.T.VV.
ITA_SA_KM247555 S.V.....V .V.QV.NM.- -DRGNTLTIS I.GNGT--D. .G.Q..E... ..A.T.VV.
Liaoning_AF327877 S.V.....K .K.KQ...R- -ENSS..IM. NCSNNSCEEF .G.S..E... ..AIT..V.
Miyazaki_JX003263 S.V.....T .QKFDADK IDNGST-TI. I.IIS.---- --.NL.E..L ..KA.T.VV.
Wyoming_AF033820 S.....Q .K.K.N.NS SDSSNSVRVE DVMN.A--E. .G....E... ..F.T..V.

```

```

BRA1      EEEMVQ-NDN NTWSPRG CNS TWARVKHCPA DLYGIHPIR LCVQPPFFLV KNNDGNFTNG
BRA2      ....I.Y...R..A.K...D ..K....M .....L.... .....RKE.NHS---
DLV9-7_HM141922 .V..Q--SK ..I.KR.KK ..K....M .....NR.. M.....F .QD.TSNNT-
DV10-3_HM141910 DI..QE-S... ..I.KR..E .R.....M .....NR.. M.....F .QD.TSNNT-
IRE_F2_JX480631 IN..RIDNS SS.V.K...E .....Q..V ..... ..DSS.DNSS
IRE_F3_JX480632 IN..RIDNS SS.V.K...E .....Q..V ..... ..DSS.DNSS
IRE_F4_JX480633 IN..I-.SS.V.K...E .....Q..V ..... ..DSS.DNSS

```

[illegible]

BRA1	NHTLSN	CGPQ	ISLGILDENK	A	VTR-EV-NC	TVAKKVFFHN	D-YSGQ	FISP	IFYQCNL	S-L
BRA2	-R....		S.	.	AVK-DA-S.	Q.E..T.Q.S	KGH...	L.V.	...K.TITN	.
DLV9-7_HM141922	-GI....	L	VF...ED.	.	AIQ-NG-S.	LHRTNIERP	.-...FYQV	...	I...TG.	.
DV10-3_HM141910	-GI....	L	VF...ED.	.	AVQ-NG-S.	LHRTNIKR	-...FYQV	...	I...TG.	.
IRE_F2_JX480631	.S.....	S	F...ED.	.	.VT-G...L	LHRNINRP	L-.Y..Y.V.	...	K.SFN-Y	.
IRE_F3_JX480632	.S.....	S	F...ED.	.	.VT-G...L	LHRNINRP	L-.Y..Y.V.	...	K.SFN-Y	.
IRE_F4_JX480633	.S.....	S	F...ED.	.	.VT-KG...LH.	LHRTNIKR	L-.Y..Y.V.	...	K.F.-Y	.
IRE_H3_JX480634	.S.....	S	F...ED.	.	.VT-G...L	LHRNINRP	L-.Y..Y.V.	...	N.F.-Y	.
ITA_DE_KM247554	.S.....	S	F...ED.	.	.VT-G-S..LHRTSIYS	L-.Y..Y.V.	...	N.SFN-Y	.	.
ITA_SA_KM247555	.S.....	S	F...ED.	.	.VT-G...L	LHRTNIKR	L-.Y..Y.V.	...	N.F.-Y	.
Liaoning_AF327877	-SI....	L	VF...ED.	.	AIQ-NG-S.	LHRTNIKR	.-...FYQV	...	I...TG.	.
Miyazaki_JX003263	-.....	VF.V.D.	D.	T	VK-SEEK.	KIIRFNIRKQ	.-.F..Y.T.	..	NYT-N-I	.
Wvoming_AF033820	-SRIG...	T	F.V.ED.	.	G.V.GDYTA.	N.NRLNINRK	.-.T.IYOV	..	T.TFTN	.

BRA1	TPCN-GTNSP	VSIIRYD	EDT VEYLIRNTVE	YLI	CKA----	---	KNITEKN	F---	SCVVQT
BRA2	.R.Y-NNS.T	I.....	.KN .Q-----	..L..T----	---	N.TDAG.	----		
DLV9-7_HM141922	QS.--NG.I	I...M.E	SNN .Q-----	..L.NT----	---	S.TNST.	NANV....	S	
DV10-3_HM141910	QS.--NG.I	I...M.E	SNN .Q-----	..L.NT----	---	S.TNST.	NANV....	S	
IRE_F2_JX480631	.S.--DD.L	I.V...E	P.N .Q-----	..L.NI----	---	N.TSDT.	NA--..I..		
IRE_F3_JX480632	.S.--ND.L	I.V.M.E	P.N .Q-----	..L.NINGTN	NTSN.TS.I.	NA--..I..			
IRE_F4_JX480633	.S.D--NDPL	I.V.M.E	P.N .Q-----	..L.NINGTN	NTSN.TRNI.	NA--..I..			
IRE_H3_JX480634	.S.D--NDPL	I.V.M.E	P.N .Q-----	..L.NINGTN	NTSN.TSKI.	NA--..I..			
ITA_DE_KM247554	.S.--ND.L	I.V.MHE	PGN LQ-----	..L.NI----	---	N.TSKI.	NA--..I..		
ITA_SA_KM247555	.S.--ND.L	I.V.M.E	P.N .Q-----	..L.NI----	---	N.TSDI.	NA--..I..		
Liaoning_AF327877	QS.--NG.I	I...MSE	SNN .Q-----	..L.NT----	---	S.TNST.	NATV....	S	
Miyazaki_JX003263	SR..QKN.KS	.V.M.E	SGN IQ-----	..L.NVNNEN	NNTNITDAIK	NY--I.IA.			
Wwoming_AF033820	.S.--NEPI	I.V.M.ETNO	.O-----	..L.NN----	---	N.SNNY.			

BRA1	FGTIGQAHIE	LP	RKNKKIIG	NQ	FTQYNC	SI	NDKTELEKWK	LVKSSGITPF	PVTSRANTGL
BRA2	.K.....	..	.N..R.RE	.K		.NQ.....	.KE..	.I.G....I	.I..Q.....
DLV9-7_HM141922	.V.....VA	...	RLQS	PN	.A....T.	.N....RQ.Q	...	T....L	.IS.T.....
DV10-3_HM141910	.V.....VA	...	RLQS	PK	.AH....	.N....RQ.Q	...	T....L	.IS.T.....
IRE_F2_JX480631	.I.....L	...	R.AT	PN	.SH...TV	.N...MQE.Q	...	T....I	.IS..NEK..
IRE_F3_JX480632	.I.....L	...	R.AT	PN	.SH...TV	.N...MQE.Q	...	T....I	.IS..NEE..
IRE_F4_JX480633	L	...	R.AT	PN	.SH...TV	.N...MQE.Q	...	T....I	.IS..NEE..
IRE_H3_JX480634	.I.....L	...	R.AT	PN	.SH...TV	.N...MQE.Q	...	T....I	.IS..NEK..
ITA_DE_KM247554	.I.....L	...	R.AE	PG	.SH...TV	.N...MQE.Q	...	T....I	.IS..SEK..
ITA_SA_KM247555	L	...	R.AE	PD	.SH...TV	.N...MQE.Q	...	T....I	.IS..NEK..
Liaoning_AF327877	.V.....VA	...	RLQS	PK	.AH....T.	.N....RQ.Q	...	T....L	.IS.T.....
Miyazaki_JX003263	.K.....L	..	.PT.G.KQ	QN	.R.....	..QG.MR..	T	T	.S.SD.S...
Wwoming_AF033820	.V.....L	..	.P..R.RN	OS	.N.....	.N....T..	...	T....L	.IS.E.....



BRA1	VRYK	DFGIS	AIVA	AIVAAT	AIAASATMSY	IALTETNKMS	DVMNHTFVVE	NDTIRATEMI
BRA2	I.....					AS.V.	NA.....E.	N.H.....
DLV9-7_HM141922	.H.....		.I.....			V.LD	G.Q....E.	N.NSM.LT
DV10-3_HM141910	.H.....		.I.....			V.LD	S.Q....E.	N.NSM.L.
IRE_F2_JX480631	.H.....L				T....	AS.IQ	.V...LN.	N.LS.....
IRE_F3_JX480632	I.H....L				T....	AS.IQ	.V...LN.	N.LS.....
IRE_F4_JX480633	I.H....L				T....	AS.IQ	.V...LN.	N.LS.....
IRE_H3_JX480634	.H....L				T....	AS.IQ	.V...LN.	N.LS.....
ITA_DE_KM247554	I.H....L				T....	AS.IQ	.V...LN.	N.LS....M
ITA_SA_KM247555	I.H....L				T....	AS.IQ	.V...LN.	N.LS.....
Liaoning_AF327877	.H.....		.I.....			V.LD	S.Q....E.	N.NGL.LV
Miyazaki_JX003263	I.....		.I.....			Q.HLM	.T....E.	N.SGMD.L
Wyoming_AF033820	I.H.....					V...V.IM	E.Q....E.	.S.LNGMDL

BRA1	EKQIHILYAM	VLQTHADVQL	LKERQKIEET	FNLIGCVERS	HTFCHTGHWP	NESWGQLNET
BRA2	. . M	I	 D . .			. D DS
DLV9-7_HM141922	. E		. . . Q	 I . .		 DS
DV10-3_HM141910	. E		. . . Q	 I . .		 DS
IRE_F2_JX480631	. R . N . . .	I	. . . K . QV . .	 I . KT		. N . . L . DS
IRE_F3_JX480632	. R . N . . .	I	. . . K . QV . .	 I . KT		. N . . F . DS
IRE_F4_JX480633	. R . N . . .	I	. . . K . QV . .	 I . T		. N . . L . DS

IRE_H3_JX480634	...VS.....	I.....	...K.QV...	I.KT		.N...L..D.
ITA_DE_KM247554	S.....	I.....	...K.QV...	I..T		.N...L..DS
ITA_SA_KM247555	.R.VS.....	I.....	...K.QV...	I..T		.N...F..DS
Liaoning_AF327877	.E.....	S	...Q.....	I.....		DS
Miyazaki_JX003263	M.....	I.....	...Q.....	.M...A...	.I.....	..T..S..DS
Wyoming_AF033820	.R..K.....	I.....	QV...	I..T	.V.....	.M...H...S

BRA1	TEWDDWITKM	NRLETEILT	LHAAKNLEQ	AMITFNTPDS	IAQFGRNIWS	HIANWIP	GLG
BRA2	.Q....VN..	.EY.QD....	R.....		K....		...
DLV9-7_HM141922	.Q....VD..	EN.NHD....	..T.R.....	S.....	K....		...
DV10-3_HM141910	.Q....VD..	EN.NHD....	..T.R.....	S.....	K....		...
IRE_F2_JX480631	.Q..T.VGE.	EKYNQD....	..I.R.....		K...D		...
IRE_F3_JX480632	.Q..T.VGE.	EKYNQD....	..I.R.....		K...D		...
IRE_F4_JX480633	.Q..T.VGE.	EKYNQD....	..V.R.....		K...D		...
IRE_H3_JX480634	.Q..T.VGE.	EKYNQD....	..V.R.....		K...D		...
ITA_DE_KM247554	.Q..T.VGE.	EKYNQD....	..I.R.....		K...D		...
ITA_SA_KM247555	.Q..T.VGE.	EKYNQD....	..I.R.....		K...D		...
Liaoning_AF327877	.Q....VD..	EN.NHD....	..T.R.....	S.....	K....		...
Miyazaki_JX003263	.Q....V.R.	.G..H.V...	R.....	SI.....T	K....	.V.....	...
Wyoming_AF033820	.Q....VS..	ED.NQ.....	..G.R...A	S.....	KDL.	..G....	...

BRA1	ASIIKYIIML	LLCYVVL	TSS	PKILRNLLRM	MSGAGSSVNR	YLKRQHHQKH	AWKEDAWDQD
BRA2	V..		TT	H	F..KK....	ES....	...
DLV9-7_HM141922	VL..	..V..L..	..A	G..T	RKRY.HR.	..SRG.I.A.V	...
DV10-3_HM141910	VL..	..V..L..	..A	G..T	RKRY.HR.	..SRG.I.A.V	...
IRE_F2_JX480631	VLF	..II....	..M	..L..H..TT	.T..A..AS.	KL.H.R	.SQDGN..R.
IRE_F3_JX480632	VLF	..II....	..M	..L..H..TT	.T....AS.	KL.H.R	.SQDGH..R.
IRE_F4_JX480633	VLF	..II....	..M	..L..H..TT	.T....AS.	KL.H.R	.SQDGN..R.
IRE_H3_JX480634	VLF	..II....	..M	..L..H..TT	.T....AS.	KL.H.R	.SQDGH..R.
ITA_DE_KM247554	VLF	..II....	..M	..L..H..TT	.T....FASH	KL.H.R	.SQDGH..R.
ITA_SA_KM247555	VLF	..II....	..M	..L..H..TT	.T....AS.	KL.H.R	.SQDGH..R.
Liaoning_AF327877	VL..	..V..L..	..A	G..T	RKRY.HR.	..SRG.I.A.V	...
Miyazaki_JX003263	..LT..V.LF	..V....	..M	..F..H..TT	I...R..AS.	...ETY.RRR	V.Q.GH....
Wyoming_AF033820	V.F	..I..LL		A.WKV	T.....GS.	...KKF.H..	..SR..T...A

BRA1	QYRIHLADAT	GGSEDKYNKP	KCSKRSYSGG	LKGFNRQPKN	WKNSTERSEV	NSTGLGEHME	
BRA2	..SV...GV.	..L.....L	.Y...N.N..	S.E..KR..S	.ETLIDA.GK	P.PS.G	
DLV9-7_HM141922	..HAY...E	H..G..S.MR	.L.RNNWN.E	SEY..RQ..	..RLIK..GE	.YN-----	
DV10-3_HM141910	..HAY...E	H..G..S.MR	.L.RNNWN.E	PEEY..RQ..	..RLIK..GE	.YNTHEDN.G	
IRE_F2_JX480631	..V...SVN	D...G.F.RQ	RAF.N.W..E	SEYSSRR.H	YEK..KEFGK	SFN..ADYE.	
IRE_F3_JX480632	..V...SVN	D...G.FSRQ	RAF.N.W..E	SEYSSRR.H	YEK..KEFGK	SFN..ADYE.	
IRE_F4_JX480633	..V...SVN	D...G.F.RQ	RAF.N.W..E	SEYSSRR.H	YEK..KEFGK	SFN..ANYE.	
IRE_H3_JX480634	..V...SVN	D...G.FSRQ	RAF.N.W..E	SEYSSRR.H	YEK..KEFGK	SFN..ANYE.	
ITA_DE_KM247554	..V...GVN	D...G.FSRQ	RAF.N.W..E	SEYSSRR.H	YEA..KKFGE	SFN..ADTK.	
ITA_SA_KM247555	..V...SVN	D...G.F.RQ	RAF.N.W..E	SEYSSRR.H	YET..KEFGE	SFN..ANNK.	
Liaoning_AF327877	..HAY...E	H..G..S.MR	.L.RNNWN.E	SEY..RQ..	..KLIK..GE	.YNTHEDN.G	
Miyazaki_JX003263	..N...NV.	V...EEF.TQ	NNFRNNWN.D	S.VYDKLQ.H	.RRL.K..GE	DWHIP.VNEG	
Wyoming_AF033820	..HN....GV.	...G...Y.Q	.Y.RNDWN.E	SEY..R..S	.VK.I.AFGE	SYISEKTKG.	

BRA1	TSHLSIAIAE	-NS-VSGDNA	QRTSTSLKIQ	SEGGNIYDCC	IKAQEGTLAI	PCCGFPLWLF	
BRA2	ITLP.M..G.	TK.-.A....	.QA.AN.E.R				
DLV9-7_HM141922	-MGRLVTT.A	EKK-NV.V.P	HQG.LN.E..			L	
DV10-3_HM141910	.MGHL.TT.A	EKK-NV.E.P	HQG.LN.E..		G.....	L	
IRE_F2_JX480631	ITPAF..PNV	KKD-AGEK.P	SQG.L..E..				
IRE_F3_JX480632	ITPAF..PNV	KKD-AGEK.P	SQG.L..E..				
IRE_F4_JX480633	ITPAF..PNV	KKD-AGEK.P	SQG.L..E..				
IRE_H3_JX480634	ITPAF..PNV	KKD-AGEK.P	SQG.L..E..				
ITA_DE_KM247554	ITPTF..PNV	KKD-AGEK.P	SQG.L..E..				
ITA_SA_KM247555	ITPTF.PPNV	KKD-AGEK.P	SQG.L..E..				
Liaoning_AF327877	.MGRLVTT.A	EKK-NV.V.P	HQG.LN.E..			L	
Miyazaki_JX003263	..HGTTTGA	KKK-HI.G.P	.QG.LD.E..	.T.....	T.		
Wyoming_AF033820	I.QPGA..N.	HKNGSG.N.P	HQG.LD.E.R				

BRA1	WGLIIIMGRL	FGYGLRGIAM	ILSYCEKGLT	MLFIMLKRLC	DFIGRALNPA	TSHVSMPOYV	
BRA2	..FV..I...	L.....	..L..G....	?.....	.S.....		
DLV9-7_HM141922	L...	L...F...K	FIMILG...N	VIITG.RK..	.Y..KM....	T....D	
DV10-3_HM141910	L...	L.....K	IMILG...N	VIITG.RK..	.Y..KM....	T....D	
IRE_F2_JX480631	L...	I.....L.K	IMFSGRVIL	LSINW.RKGL	.YV.KM...S	K...T..C.D	
IRE_F3_JX480632	L...	I.....L.K	IMFSGRVIL	LSINW.RKGL	.YV.KM...S	K...T..C.D	
IRE_F4_JX480633	L...	I.....L.K	IMFSGRVIL	LSINW.RKGL	.YV.KM...S	K...T..C.D	

```

IRE_H3_JX480634      . . . . . L . . . I . . . F . . L . K . IMFSGRVIL LSINW.RKGL .YV.KM. . . S K . . . T . . C . D
ITA_DE_KM247554      . . . . . L . . . I . . . . . L . K . IMFSGRVIL LSINW.RKGL .YV.KM. . . S K . . . T . . C . D
ITA_SA_KM247555      . . . . . L . . . I . . . . . L . K . IMFSGRVIL LSINW.RKGL .YV.KM. . . S K . . . T . . C . D
Liaoning_AF327877    . . . . . L . . . L . . . . . . . K . IMILG . . . N VIITG.RK . . . Y . . KM . . . . . T . . . . D
Miyazaki_JX003263    . . . L . . . . . L . . . . K . L . I M . QIIG . . . Y G . LN . I . QIF . YM . N . FS . P K . . I . . . . I
Wyoming_AF033820     . . . V . . V . . I A . . . . . L . V . IRI . IR . . N LI . EIIRKML . Y . . . . . G . . . . . . .

```

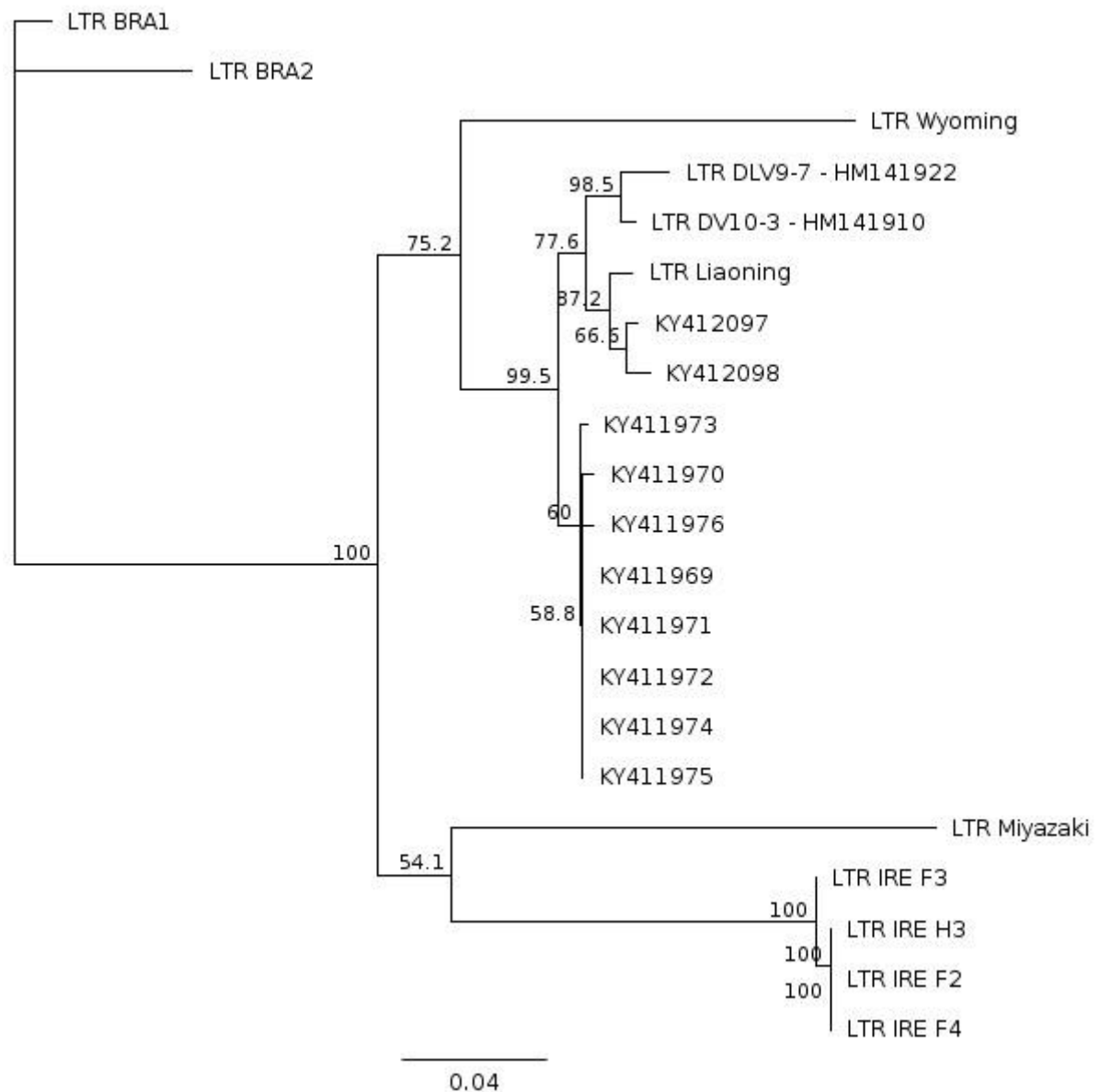
```

BRA1      --
BRA2      --
DLV9-7_HM141922  V-
DV10-3_HM141910  V-
IRE_F2_JX480631  I-
IRE_F3_JX480632  I-
IRE_F4_JX480633  I-
IRE_H3_JX480634  I-
ITA_DE_KM247554  I-
ITA_SA_KM247555  I-
Liaoning_AF327877  V?
Miyazaki_JX003263  --
Wyoming_AF033820  --

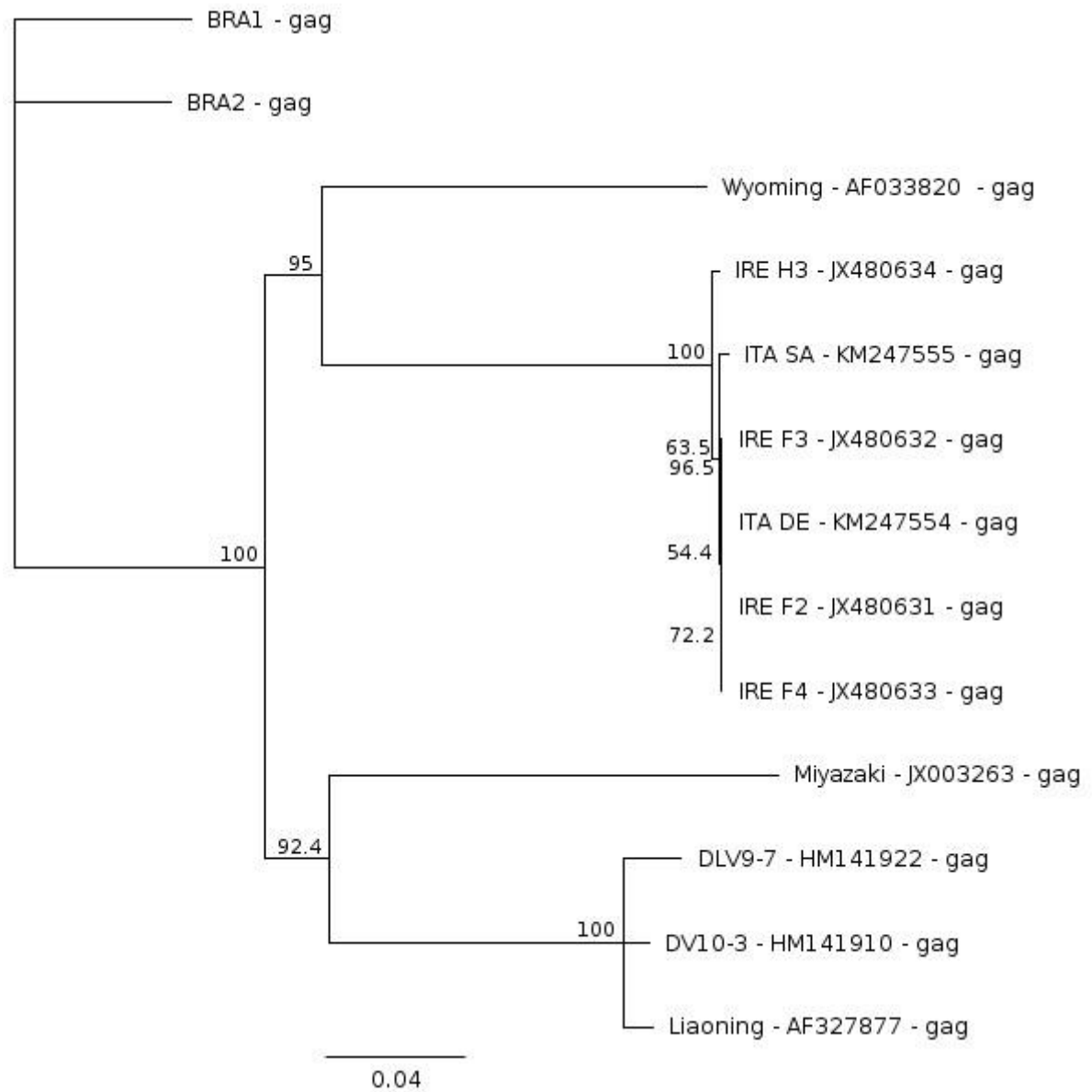
```

Appendix G: Phylogenetic trees of Brazilian field sequences BRA1 and BRA2 and foreign sequences. Each tree was performed based a genome region or gene: A. LTR; B. Gag; C. Pol; D. Env; E. Env protein. Used sequences: Liaoning (AF327877), Wyoming (AF033820), Miyazaki 2011-A (JX003263), IRE F2 (JX480631) IRE F3 (JX480632), IRE F4 (JX480633), IRE H3 (JX480634), ITA SA (KM247555), ITA DE (KM247554), DLV9-7 (HM141922), DV10-3 (HM141910), and LTR sequences from China (KY411969-76; KY412097-98).

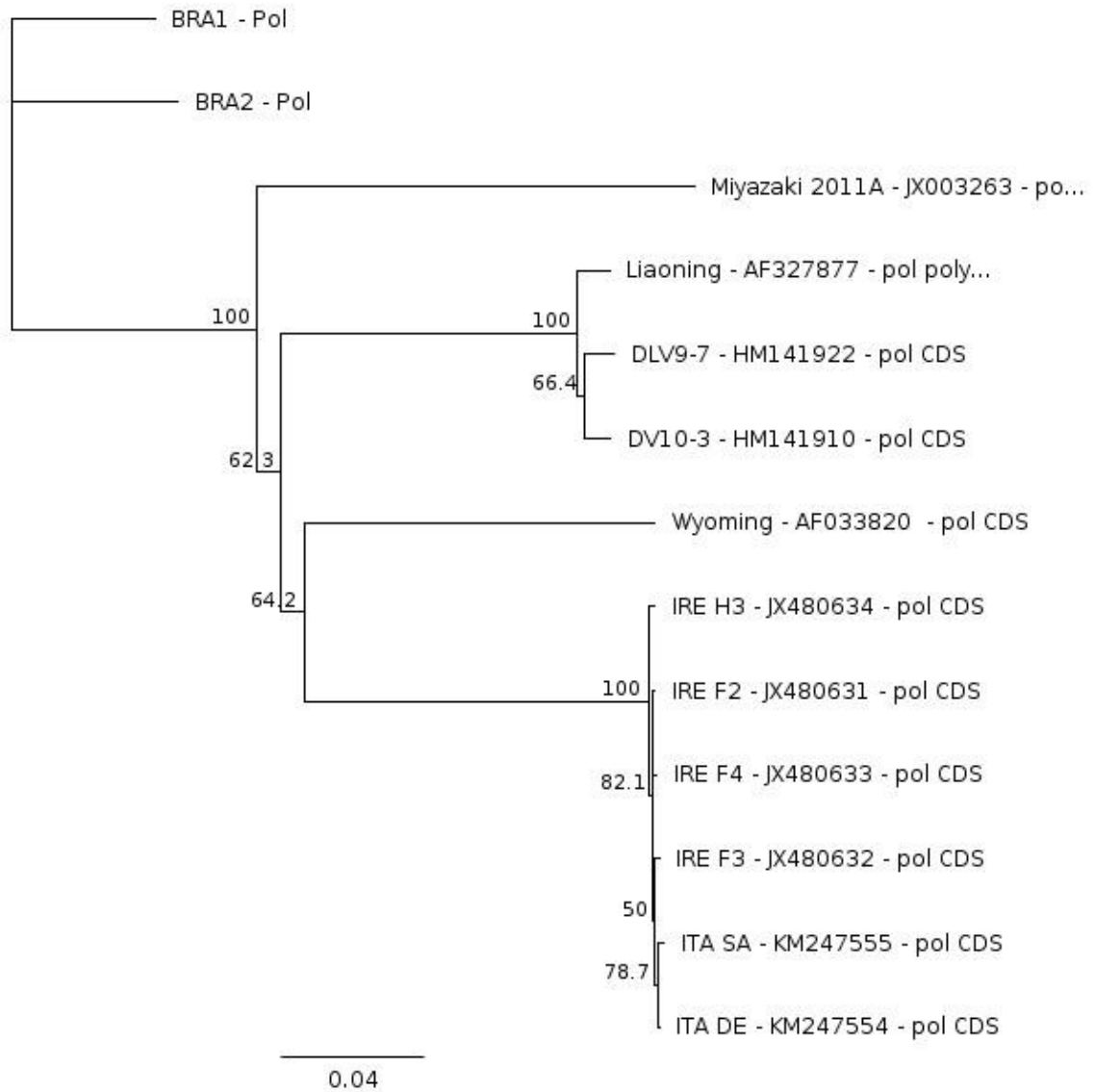
A.



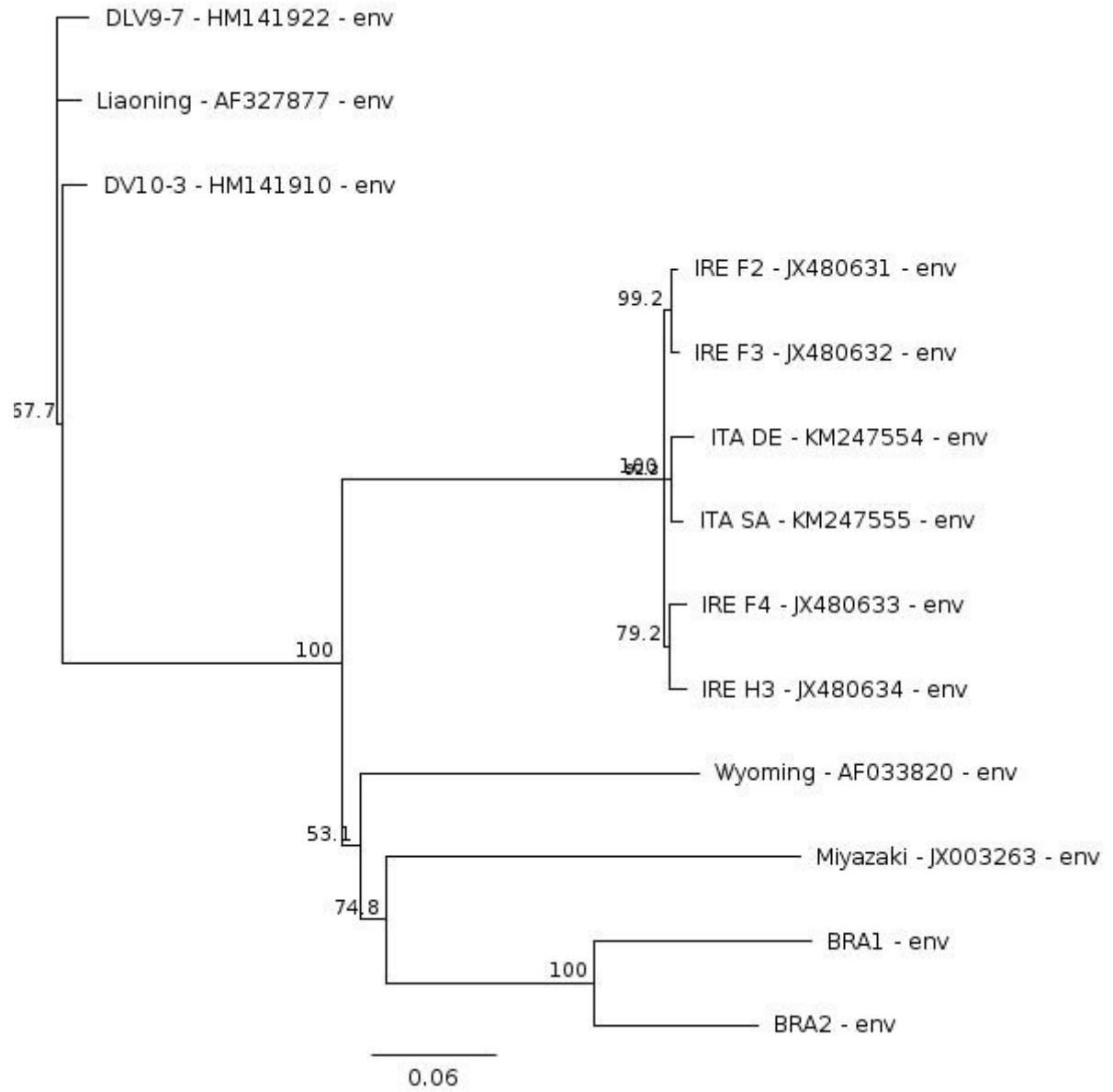
B.



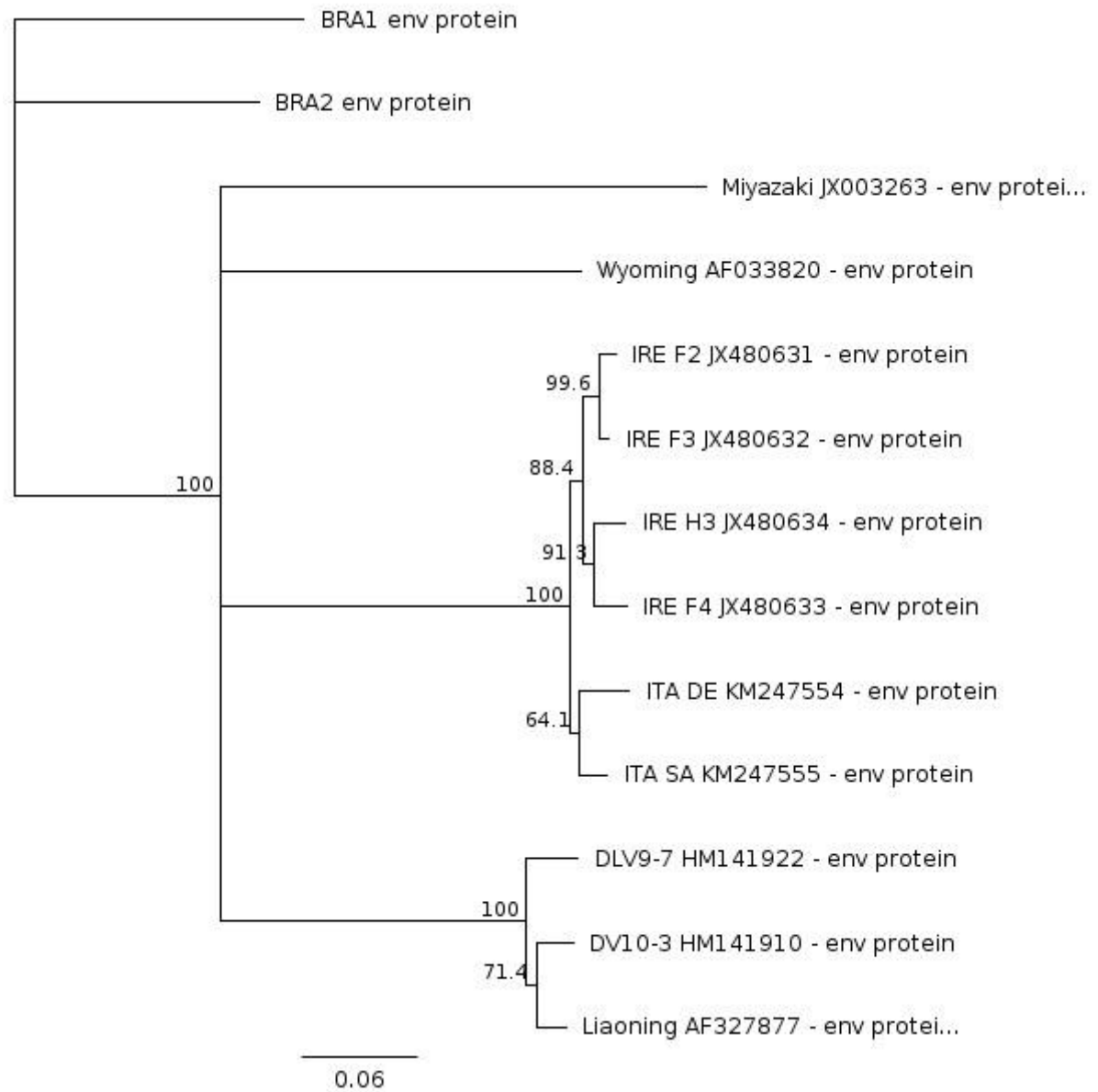
C.



D.



E.



Appendix H: Molecular modeling of BRA1 and BRA2 p26 using a p26 crystallographic structure (PDB code 1EIA) with Modeller software. Epitopes with amino acids alterations were compared in both models, showing surface amino acids alterations. Blue arrows point to amino acid alteration compared to the reference.

