RAFAEL DE NOVAES OLIVEIRA

**Modos e tempo de evolução em linhagens do vírus da raiva (RABV) mantidos por
reservatórios aéreos e terrestres com base em genomas completos**

São Paulo

2014

RAFAEL DE NOVAES OLIVEIRA

Modos e tempo de evolução em linhagens do vírus da raiva (RABV) mantidos por reservatórios aéreos e terrestres com base em genomas completos

Tese apresentada ao Programa de Pós-Graduação em Epidemiologia Experimental Aplicada às Zoonoses da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo para obtenção do título de Doutor em Ciências

Departamento:

Medicina Veterinária Preventiva e Saúde Animal

Área de concentração:

Epidemiologia Experimental Aplicada às Zoonoses

Orientador:

Prof. Dr. Paulo Eduardo Brandão

De acordo: _____

Orientador

São Paulo
2014

Obs: A versão original se encontra disponível na Biblioteca da FMVZ/USP

Autorizo a reprodução parcial ou total desta obra, para fins acadêmicos, desde que citada a fonte.

DADOS INTERNACIONAIS DE CATALOGAÇÃO-NA-PUBLICAÇÃO

(Biblioteca Virginie Buff D'Ápice da Faculdade de Medicina Veterinária e Zootecnia da
Universidade de São Paulo)

T.2986
FMVZ

Oliveira, Rafael de Novaes

Modos e tempo de evolução em linhagens do vírus da raiva (RABV) mantidos por reservatórios aéreos e terrestres com base em genomas completos / Rafael de Novaes Oliveira. -- 2014.

135 p. : il.

Tese (Doutorado) - Universidade de São Paulo. Faculdade de Medicina Veterinária e Zootecnia. Departamento de Medicina Veterinária Preventiva e Saúde Animal, São Paulo, 2014.

Programa de Pós-Graduação: Epidemiologia Experimental Aplicada às Zoonoses.

Área de concentração: Epidemiologia Experimental Aplicada às Zoonoses.

Orientador: Prof. Dr. Paulo Eduardo Brandão.

1. Raiva. 2. Vírus. 3. Evolução. 4. Chiroptera. 5. Carnívora. I. Título.

UNIVERSIDADE DE SÃO PAULO

FACULDADE DE MEDICINA VETERINÁRIA E ZOOTECNIA

Comissão de Ética no uso de animais

CERTIFICADO

Certificamos que o Projeto intitulado "Modos e tempo de evolução em linhagens do vírus da raiva (RABV) mantidos por reservatórios aéreos e terrestres com base em genomas completos", protocolado sob o nº 2030/2010, utilizando 30 (trinta) camundongos, sob a responsabilidade do Prof. Dr. Paulo Eduardo Brandão, está de acordo com os princípios éticos de experimentação animal da "Comissão de Ética no uso de animais" da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo e foi aprovado em reunião de 29/09/2010.

We certify that the Research "Modes and duration in strains of rabies (RABV) maintained by air and land reservoirs viruses based on complete genomes", protocol number 2030/2010, utilizing 30 (thirty) mice, under the responsibility Prof. Dr. Paulo Eduardo Brandão, agree with Ethical Principles in Animal Research adopted by "Ethic Committee in the use of animals" of the School of Veterinary Medicine and Animal Science of University of São Paulo and was approved in the meeting of day 09/29/2010.

São Paulo, 28 de maio de 2014

Prof.ª Dra. Alice Maria Melville Paiva Della Libera
Vice-Presidente no exercício

Av. Prof. Dr. Orlando Marques de Paiva, nº87
Cidade Universitária "Armando de Salles Oliveira"
São Paulo/SP – Brasil
05508-270

Fone: + 55 11 3091-7671/7676/0904/8359
Fax: +55 11 3032-2224
E-mail: fmvz@usp.br
<http://www.fmvz.usp.br>

FOLHA DE AVALIAÇÃO

Autor: OLIVEIRA, Rafael de Novaes

Título: Modos e tempo de evolução em linhagens do vírus da raiva (RABV) mantidos por reservatórios aéreos e terrestres com base em genomas completos

Tese apresentada ao Programa de Pós-Graduação em Epidemiologia Experimental Aplicada às Zoonoses da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo para obtenção do título de Doutor em Ciências

Data: ____/____/____

Banca Examinadora

Prof. Dr. _____

Instituição: _____ Julgamento: _____

Prof. Dr. _____

Instituição: _____ Julgamento: _____

Prof. Dr. _____

Instituição: _____ Julgamento: _____

Prof. Dr. _____

Instituição: _____ Julgamento: _____

Prof. Dr. _____

Instituição: _____ Julgamento: _____

"O que é isso que lhe chamou tão repentinamente do nada para desfrutar por um breve tempo um espetáculo que permanece bastante indiferente a você? As condições para sua existência são tão antigas quanto às pedras. Por milhares de anos, os homens se esforçaram, sofreram e procriaram e as mulheres deram à luz em dor. Cem anos atrás, talvez, outro homem – ou mulher – se sentou neste ponto; como você, ele contemplou, com temor e receio no coração, o morrer das geleiras. Como você, ele foi criado de um homem e nascido de uma mulher. Ele sentia dor e breve alegria, como você sente. Ele era outra pessoa? Não era ele você mesmo?"

Erwin Schrödinger

Somos todos animais com roupas

Sibele Pinheiro de Souza

AGRADECIMENTOS

Ao Prof. Dr. Paulo Eduardo Brandão, pela amizade e orientação.

A toda a minha família, principalmente meus pais Esther de Novaes Oliveira, Vagner Moraes de Oliveira, meus irmãos Fernando de Novaes Oliveira, Mariana de Novaes Oliveira e Pedro de Novaes Oliveira, pelo apoio e compreensão.

A minha companheira de vida Fabiana Schiavi Noda, pelo companheirismo e paciência.

A grande amiga Sibele Pinheiro de Souza, por tudo mesmo

A Juliana Levino por salvar minha pele.

Aos colegas e amigos do VPS, em especial Anaiá, Ekaterina e Vanessa.

Aos amigos e colegas do instituto pasteur, em especial a Dra. Luciana (Diretora) Samira, Pedro, Juliana, Helena, Carla, Andréa, Bilu e Graciane, Geralda, Dra. Zélia, Cidoka, e todos os demais colegas.

Ao pessoal do LEMB no ICB2, por me acolherem tão bem e me ajudarem em tudo que eu precisei, em especial ao professor Paolo Zanotto, Caio, Carla e Átila.

Aos amigos de sempre, Betinho, Everton, Mateus, Chanel e Lucas.

Ao Shimbau, o melhor amigo que tive.

RESUMO

OLIVEIRA, R. N. **Modos e tempo de evolução em linhagens do vírus da raiva (RABV) mantidos por reservatórios aéreos e terrestres com base em genomas completos.** [Modes and time of evolution of Rabies virus (RABV) lineages found in aerial and terrestrial reservoirs based on complete genomes]. 2014. 135 f. Tese (Doutorado em ciências) - Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2014.

A raiva é uma zoonose que afeta o sistema nervoso central, de evolução aguda e fatal, mantida em mamíferos e conhecida há milênios. Presente na América, Europa, África e Ásia, tem como agente etiológico o vírus da raiva (RABV), um vírus RNA neurotrópico, pertencente à ordem *Mononegavirales*, família *Rhabdoviridae*, gênero *Lyssavirus*., o qual é composto por quatorze espécies. Entre os *Lyssavirus*, o RABV é o mais amplamente distribuído mundialmente e tem maior importância epidemiológica dada sua associação com um maior número de casos de encefalite por *Lyssavirus* em humanos em relação às outras espécies. São admitidos dois ciclos de transmissão para a raiva, o ciclo urbano e o ciclo silvestre. O ciclo urbano ou terrestre tem o cão como principal reservatório e transmissor do vírus para outros cães, outros animais domésticos e para o homem, enquanto o ciclo silvestre ou aéreo é mantido por diferentes mamíferos silvestres e quirópteros. A origem comum dos dois ciclos do RABV a partir de um RABV ou *Lyssavirus* ancestral e a divergência adaptativa ocorrida desde então, causada pela adaptação de tal vírus em paisagens adaptativas tão variadas e distintas representadas pelas ordens *Carnivora* e *Chiroptera*, levaram ao surgimento das diversas linhagens encontradas nos ciclos terrestre e aéreo. Sendo assim, com o objetivo de se estudar as diferenças geradas nos RABV dos ciclos aéreo e terrestres devido a sua evolução em paralelo nestas duas ordens foram analisadas 159 sequências genômicas do RABV (59 do ciclo terrestre e 100 do ciclo aéreo), sendo que 21 destas sequências foram obtidas neste estudo e representam oito linhagens de RABV existentes no Brasil e cinco destas linhagens de RABV tiveram seus genomas sequenciados pela primeira vez. Foram analisados aspectos como as diferentes taxas de substituição de nucleotídeos por sítios (heterotaquia) entre os mesmos genes do RABV mantidos no ciclo aéreo e terrestre, análise do melhor gene para a realização de estudo filogenéticos confiáveis para o RABV, tempo de divergência entre os ciclos, padrões de variabilidade genética e vieses quanto ao uso preferencial de códons em cada ciclo. Como resultado, concluí-se que a divergência

adaptativa ocorrida entre os dois ciclos do RABV fez com que alguns aspectos evolutivos de seu genoma apresentem padrões diferentes de acordo com o ciclo do RABV analisado.

Palavras-chave: Raiva. Vírus. Evolução. *Chiroptera*. *Carnivora*.

ABSTRACT

OLIVEIRA, R. N. **Modes and time of evolution of Rabies virus (RABV) lineages found in aerial and terrestrial reservoirs based on complete genomes** [Modos e tempo de evolução em linhagens do vírus da raiva (RABV) mantidos por reservatórios aéreos e terrestres com base em genomas completos]. 2014. 135 f. Tese (Doutorado) - Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2014.

Rabies is a zoonosis that affects the central nervous system, showing an acute and fatal evolution, occurring in mammals and known for millennia. Present in America, Europe, Africa and Asia, its etiological agent is Rabies virus (RABV), a neurotropic RNA virus in the order *Mononegavirales*, family *Rhabdoviridae*, genus *Lyssavirus*, composed by fourteen species. Amongst the lyssaviruses, RABV is the most widely spread worldwide and has a higher epidemiological importance due to its association to a higher number of cases of encephalitis. Two cycles are accepted for rabies transmission, the urban and the wild ones. In the urban (or terrestrial) cycle, dogs are the main reservoirs and transmitters of the virus to other dogs, other domestic animals and to humans, while in the wild (or aerial) cycle bats are the reservoirs. The common origin of both cycles from an ancestor RABV or lyssavirus and the adaptive divergence that occurred since then, caused by the adaptation of this ancestor virus to a wide range of adaptive landscapes represented by the orders *Carnivora* and *Chiroptera* led to the emergence of diverse RABV lineages currently found in the aerial and terrestrial cycles. Thus, aiming to study differences found in RABV lineages from the aerial and terrestrial cycles due to their parallel evolution in these two orders, 159 genomic sequences of RABV (59 from the terrestrial and 100 from the aerial cycles) were analyzed, being 21 of these sequences referent to eight lineages of RABV found in Brazil sequenced in this study and five of these eight lineages of RABV had their genomes sequenced for the first time. The study included the per site nucleotide substitution rate differences (heterotachy) between the same genes RABV maintained in the aerial and terrestrial, survey of the most suitable gene for phylogenetic analysis, time of divergence between the two cycles, patterns of genetic variability and codon usage bias. As a conclusion, the adaptive divergence occurred between the two cycles caused some evolutionary aspects of RABV genome to show an intricate cycle-specific evolutionary pattern.

Keywords: Rabies. Virus. Evolution. *Chiroptera*. *Carnivora*.

LISTA DE ABREVIATURAS E SÍMBOLOS

ICTV	International Committee on Taxonomy of Viruses
%	Porcento
BLAST/n	Basic Local Alignment Search Tool
RNA	ácido ribonucleico
RABV	Rabies virus
DUVV	Duvenhage virus
EBLV-1	European bat lyssavirus – 1
EBLV-2	European bat lyssavirus – 2
ABLV	Australian bat lyssavirus
LBV	Lagos bat vírus
MOKV	Mokola vírus
AcM	Anticorpo monoclonal ou Anticorpos monoclonais
RT	Transcrição Reversa
PCR	Reação em Cadeia pela Polimerase
WCBV	West Caucasian bat vírus
RNP	Ribonucleoproteína
N	Nucleoproteína do vírus da raiva
L	RNA
P	Fosfoproteína do vírus da raiva
G	Glicoproteína do vírus da raiva
RT	Transcrição reversa
M	Molar
M	Proteína M do vírus da raiva
mRNA	RNA mensageiro
Ψ	Pseudo gene
pb	pares de bases
L	Proteína L do vírus da raiva
Le	Seqüência leader
Tr	Seqüência trailer
SNC	Sistema nervoso central
IFD	Imunofluorescência direta
CVS	Challenger Virus Standard
N2A	Neuroblastoma de camundonto
dNTP	deoxinucleosídeo-trifosfato
pmol	Picomoles
mM	Milimolar
U	unidade internacional
μL	Microlitro
RNAsin	Inibidor de RNase
H ₂ O	Água
cDNA	DNA complementar
TBE	Tampão Tris borato
ng	Nanograma
°C	graus Celsius
BLAST/n	Basic Local Alignment Search Tool
VSV	Virus da estomatite vesicular
pH	Potencial hidrogeniônico
DNA	ácido desoxirribonucléico
EUA	Estados Unidos da América
Min	minuto de hora
%	Porcento
kDa	QuiloDalton
M	Molar
ng	nanogramas

mL	mililitro
µg	micrograma
pM	picomolar
ORF	open reading frame

LISTA DE ILUSTRAÇÕES

Quadro 1 – Relação das variantes antigênicas do RABV encontradas nas Américas e seus respectivos reservatórios - São Paulo - 2014.....	29
Quadro 2 – Relação das amostras utilizadas nesta pesquisa - São Paulo 2014.....	48
Quadro 3 – Primers utilizados para a realização das reações de Transcrição Reversa do genoma completo do RABV, PCR controle, PCR do genoma completo do RABV e sequenciamento de DNA de suas extremidades 3' e 5' UTR.-São Paulo – 2014.....	50
Quadro 4 – Ciclos de temperaturas utilizados na PCR controle da obtenção do c-DNA genômico dos RABV testados – São Paulo – 2014	51
Quadro 5 – Ciclos de temperaturas utilizados na PCR para a amplificação do genoma do RABV da amostra CVS-11 – São Paulo – 2014	52
Quadro 6 – Tamanho das sequências de DNA obtidas e média de leituras de cada nucleotídeo para cada amostra - São Paulo - 2014.....	69
Quadro 7 – Mutações encontradas entre as amostras CVS-11 GQ918139.1 e a amostra CVS-11 Pasteur - São Paulo - 2014	72
Quadro 8 – Taxas de substituição de nucleotídeos por sítio e TACMR encontrados para os genes N, P, M, G, L separadamente nos ciclos aéreos e terrestres e para os dois ciclos unidos utilizando-se os genes G+L concatenados - São Paulo – 2014.....	74
Quadro 9 – Descrição do sinal filogenético encontrado para cada gene e genoma do RABV - São Paulo - 2014	77
Quadro 10 – Aminoácidos e respectivos códons para posição 193 da proteína G do ciclo terrestre - São Paulo - 2014.....	78
Quadro 11 – Número de sítios sob seleção purificadora nos 5 genes do RABV, porcentagem de sítios sob seleção purificadora em relação ao número de aminoácidos total de cada proteína e respectivos valores de p (teste exato de Fisher) - São Paulo - 2014.....	78
Quadro 12 – Valores de ω (dN-dS) para os cinco genes do RABV calculados separadamente para os ciclos aéreo e terrestre - São Paulo - 2014.....	79
Quadro 13 – Índice de Adaptação de Códons encontrado para cada gene do RABV nos ciclos aéreo e terrestre - São Paulo - 2014.....	80

Quadro 14 – Número efetivo de códons utilizados (Nc) para cada gene do RABV nos ciclos aéreo e terrestre - São Paulo - 2014	82
Quadro 15 – Conteúdo GC3 encontrado para cada gene do RABV nos ciclos aéreo e terrestre - São Paulo - 2014.....	82
Quadro 16 – Distribuição por amostra dos códons utilizados preferencialmente em cada gene do RABV nos ciclos aéreo e terrestre - São Paulo - 2014	84
Quadro 17 – Valores medianos, mínimos e máximos encontrados para os valores de distância patrística nos ciclos aéreo e terrestre do RABV analisados separadamente e em conjunto Paulo - 2014.....	85
Quadro 18 – Variabilidade genética encontrada para os 5 genes do RABV calculada através dos métodos de distância patrística, identidade genética e distância genética, analisadas para os ciclos aéreo e terrestre do RABV- São Paulo - 2014.....	86
Quadro 19 – Genes mais conservados nos ciclos aéreo e terrestre do RABV, analisados separadamente e em conjunto, comparando os resultados de distância patrística, distância genética e identidade genética - São Paulo - 2014	87
Quadro 20 – Genes mais variáveis nos ciclos aéreo e terrestre do RABV, analisados separadamente e em conjunto, comparando os resultados de distância patrística, distância genética e identidade genética - São Paulo - 2014	87
Quadro 21 – Coeficiente de correlação para as divergências genéticas encontradas para os cinco genes do RABV no ciclo aéreo através dos métodos de distância patrística, distância genética e identidade genética - São Paulo - 2014	87
Quadro 22 – Coeficiente de correlação para as divergências genéticas encontradas para os cinco genes do RABV no ciclo terrestre através dos métodos de distância patrística, distância genética e identidade genética - São Paulo - 2014.....	88
Quadro 23 – Coeficiente de correlação para as divergências genéticas encontradas para os cinco genes do RABV através dos métodos de distância patrística, distância genética e identidade genética - São Paulo - 2014.....	89
Figura 1 – Visualização em gel de agarose a 1% da banda de aproximadamente 12 kpb resultante da amplificação do genoma completo da amostra CVS-11 - São Paulo - 2014.....	67
Figura 2 – Mutação de nucleotídeo e substituição do último aminoácido da nucleoproteína da amostra IP 1770/12 - São Paulo - 2014	70
Figura 3 – Inserção de adenina na posição 5332 da amostra IP 1770/12 (linhagem <i>C. jacchus</i>) em relação a amostras da mesma linhagem, entre os nt 5363-5364 em relação a amostra PV M13215 - São Paulo - 2014.....	70

Figura 4 –	Inserção de timina na posição 2430 amostra IP 4005/12 (linhagem <i>D. rotundus</i>) em relação a amostras da mesma linhagem na posição 2452 da amostra PV M13215 - São Paulo - 2014.....	70
Figura 5 –	Inserção de adenina na posição 5334 da amostra IP 4005/12 (linhagem <i>D. rotundus</i>) em relação a amostras da mesma linhagem, entre os nucleotídeos 5363-5364 da amostra PV M13215 - São Paulo - 2014.....	71
Figura 6 –	Inserção de adenina na posição 2454 da amostra IP 1400/10 (linhagem <i>Myotis</i>) em relação a amostras da mesma linhagem, entre os nucleotídeos 2475-2476 da amostra PV M13215 - São Paulo - 2014.....	71
Figura 7 –	Árvores com máxima credibilidade de clados. A árvore foi inferida a partir dos genes G e L concatenados. Os ramos estão calibrados pelas datas de isolamento em anos. Na parte inferior está representada a escala dos eventos de separação entre as linhagens de RABV. Os grupos foram coloridos de acordo com a classificação proposta por Bourhy et al., 2008 e Kuzmin et al., 2012 - São Paulo - 2014.....	75
Figura 8 –	Árvore filogenética inferida por máxima verossimilhança a partir de sequências genômicas. Os grupos foram coloridos de acordo com a classificação proposta por Bourhy et al., 2008 e Kuzmin et al., 2012 - São Paulo - 2014.....	76
Figura 9 –	Intensidade da seleção purificadora ($\omega < 0$) atuante nos 5 genes do RABV e os respectivos valores de p - São Paulo - 2014.....	79
Figura 10 –	Índice de Adaptação de Códon para os 5 genes do RABV analisados separadamente para os ciclos aéreo (azul) e terrestre (marrom) com os respectivos valores de p para a significância estatística das diferenças encontradas - São Paulo - 2014.....	81
Figura 11 –	A) boxplot mostrando o Nc de cada gene do RABV separado em ciclo aéreo (azul) e ciclo terrestre (marro) com os respectivos valores de p para a significância estatística para as diferenças encontradas entre os dois ciclos; B a F) cinco gráficos mostrando o esperado (curva no gráfico) e o observado (pontos no gráfico) efetivo número de códons (Enc e Nc respectivamente) em relação ao conteúdo GC3 observado para os genes N, P, M, G e L do RABV nos ciclos aéreo (azul) e terrestre (marrom) do RABV - São Paulo - 2014.....	83
Figura 12 –	Estimativa da distância patrística para os genes N, P, M, G e L encontradas nos ciclos aéreo e terrestre do RABV- São Paulo - 2014	85
Figura 13 –	Distâncias patrísticas encontradas para os cinco genes do RABV analisando os ciclos aéreo e terrestre conjuntamente - São Paulo - 2014.....	89

SUMÁRIO

1	INTRODUÇÃO	25
2	OBJETIVOS	45
3	MATERIAL E MÉTODOS	47
3.1	AMOSTRAS	47
3.2	SÍNTESE DE DNA COMPLEMENTAR (cDNA) DE TAMANHO GENOMICO	49
3.2.1	Extração de RNA.....	49
3.3	TRANSCRIÇÃO REVERSA (RT) PARA A OBTENÇÃO DO DNA COMPLEMENTAR (cDNA) AO ANTI-GENOMA COMPLETO DO RABV	49
3.4	REAÇÃO EM CADEIA PELA POLIMERASE (PCR) DIRIGIDA A EXTREMIDADE 3' UTR DO RABV - PCR-CONTROLE.	50
3.5	PCR DE AMPLICON DE TAMANHO GENOMICO.....	52
3.6	AMPLIFICAÇÃO DO GENOMA DO RABV EM DUAS REAÇÕES DE PCR.....	53
3.7	PURIFICAÇÃO DAS REAÇÕES DE PCR	53
3.8	SEQUENCIAMENTO DE DNA PARA A CONFIRMAÇÃO DA AMPLIFICAÇÃO DO GENOMA COMPLETO DO RABV E DAS REAÇÕES DE PCR A E B.....	53
3.9	PREPARO DA BIBLIOTECA DE DNA DAS AMOSTRAS E SEQUENCIAMENTO GENOMICO POR NEXT GENERATION SEQUENCING (NGS) NO SEQUENCIADOR ILLUMINA MISEQ	54
3.9.1	Diluição das amostras	54
3.9.2	Fragmentação, inserção dos adaptadores e neutralização	55
3.9.3	Amplificação	55
3.9.4	Purificação	56
3.9.5	Normalização	56
3.10	AMPLIFICAÇÃO DOS CLUSTERS E SEQUENCIAMENTO POR SÍNTESE NO SEQUENCIADOR ILLUMINA MISEQ.....	57

3.11	EDIÇÃO DE SEQUÊNCIAS	57
3.12	ANÁLISE DESCRITIVA DAS SEQUÊNCIAS OBTIDAS NESTA PESQUISA ..	58
3.13	CONJUNTO DE DADOS PARA AS ANÁLISES	58
3.14	ANÁLISE DE RECOMBINAÇÃO.....	59
3.15	ANÁLISE DE SINAL FILOGENÉTICO	59
3.16	ANÁLISE DE HETEROTAQUIA	60
3.17	ESTIMATIVAS DA TAXA DE SUBSTITUIÇÃO DE NUCLEOTÍDEOS POR SÍTIO, TEMPO PARA O ANCESTRAL COMUM MAIS RECENTE (TACMR) E ESTIMATIVAS DE FILOGENIAS	61
3.17.1	Estimativa da taxa de substituição de nucleotídeos por sítio e tempo para o ancestral comum mais recente (TACMR) calculadas separadamente para os ciclos aéreo e terrestre do RABV.....	61
3.17.2	Estimativa da taxa de substituição de nucleotídeos e tempo para o ancestral comum mais recente (TACMR) calculadas em conjunto para os ciclos aéreo e terrestre do RABV	62
3.18	ANÁLISE DO REGIME DE SELEÇÃO	63
3.19	ANÁLISE DO USO PREFERENCIAL DE CÓDONS	64
3.20	DIVERSIDADE GENÉTICA CALCULADA ATRAVÉS DA DISTÂNCIA PATRÍSTICA E SUA COMPARAÇÃO COM OS RESULTADOS DE DISTÂNCIA GENÉTICA E IDENTIDADE GENÉTICA.....	65
3.21	TESTES ESTATÍSTICOS.....	66
4	RESULTADOS	67
4.1	RT- PCR PARA AMPLIFICAÇÃO DO GENOMA COMPLETO DO RABV	67
4.2	AMPLIFICAÇÃO DO GENOMA DO RABV EM DUAS REAÇÕES DE PCR	68
4.3	SEQUENCIAMENTO DE DNA PARA A CONFIRMAÇÃO DA AMPLIFICAÇÃO DO GENOMA COMPLETO DOS RABV AMPLIFICADOS E DAS REAÇÕES DE PCR A E B	68
4.4	AMPLIFICAÇÃO DOS CLUSTERS E SEQUENCIAMENTOS POR SÍNTESE NA PLATAFORMA ILLUMINA MISEQ.	68
4.5	ANÁLISE DESCRITIVA DAS SEQUÊNCIAS OBTIDAS NESTA PESQUISA. .	70
4.6	ANÁLISE DE RECOMBINAÇÃO.....	72

4.7	ANÁLISE DE HETEROTAQUIA	73
4.8	ESTIMATIVAS DA TAXA DE SUBSTITUIÇÃO DE NUCLEOTÍDEOS POR SÍTIO, TEMPO PARA O ANCESTRAL COMUM MAIS RECENTE (TACMR) E FILOGENIAS	73
4.9	ANÁLISE DO SINAL FILOGENÉTICO	77
4.10	ANÁLISE DO REGIME DE SELEÇÃO	77
4.11	ANÁLISE DO USO PREFERENCIAL DE CÓDONS	80
4.11.1	Índice de Adaptação de Códon (CAI)	80
4.11.2	Número efetivo de códon utilizados (Nc) e conteúdo GC3.....	81
4.11.3	Uso relativo de códon sinônimos (RSCU)	83
4.12	DIVERSIDADE GENÉTICA CALCULADA ATRAVÉS DA DISTÂNCIA PATRÍSTICA E SUA COMPARAÇÃO COM OS RESULTADO DE DISTÂNCIA GENÉTICA E IDENTIDADE GENÉTICA.	85
4.12.1	Análise em separado dos ciclos aéreo e terrestre do RABV	85
4.12.2	Análise em conjunto dos ciclos aéreo e terrestre do RABV	88
5	DISCUSSÃO	91
5.1	PADRONIZAÇÃO DA REAÇÃO DE RT-PCR GENÔMICA DO RABV E SUA UTILIZAÇÃO PARA A REALIZAÇÃO DO SEQUENCIAMENTO DE DNA POR AMPLIFICAÇÃO NA PLATAFORMA ILLUMINA MISEQ.....	91
5.2	ANÁLISE DESCRITIVA DAS LINHAGENS ESTUDADAS	93
5.3	ANÁLISES DE RECOMBINAÇÃO.....	96
5.4	ANÁLISE DE HETEROTAQUIA	97
5.5	ESTIMATIVAS DA TAXA DE SUBSTITUIÇÃO DE NUCLEOTÍDEOS POR SÍTIO E TEMPO DE ANCESTRAL COMUM MAIS RECENTE (TACMR) E ESTIMATIVA DE FILOGENIAS.....	99
5.6	ANÁLISE DO SINAL FILOGENÉTICO	107
5.7	ANÁLISE DO REGIME DE SELEÇÃO	108
5.8	ANÁLISE DO USO PREFERENCIAL DE CÓDONS	115
5.8.1	Índice de Adaptação de Códon.....	115
5.8.2	Número efetivo de códon utilizados (Nc) e conteúdo GC3.....	116

5.8.3	Uso relativo de códons sinônimos (RSCU)	118
5.9	DIVERSIDADE GENÉTICA CALCULADA ATRAVÉS DA DISTÂNCIA PATRÍSTICA E SUA COMPARAÇÃO COM OS RESULTADOS DE DISTÂNCIA GENÉTICA E IDENTIDADE GENÉTICA.....	119
6	CONCLUSÕES.....	123
	REFERÊNCIAS.....	127

1 INTRODUÇÃO

A raiva é uma zoonose que afeta o sistema nervoso central, de evolução aguda e fatal, mantida em mamíferos e conhecida há milênios. Presente na América, Europa, África e Ásia, tem como agente etiológico um vírus RNA neurotrópico, pertencente à ordem *Mononegavirales*, família *Rhabdoviridae*, gênero *Lyssavirus* (WUNNER, 2007; KING et al., 2012).

Apesar dos grandes avanços ocorridos no desenvolvimento de vacinas e medidas de controle contra a raiva, estima-se que entre 40 a 70 mil pessoas morram anualmente devido a doença. Além disso, pesquisas mostram que este número pode ser até 100 vezes maior e que por volta de apenas 3% dos casos de raiva são relatados oficialmente (SCHNELL et al., 2010).

Os *Lyssavirus* são vírus envelopados, com o formato de “bala de revólver”, apresentando-se como quatorze genótipos, agrupados em dois filogrupos, genética e imunopatologicamente distintos (TORDO, 1996; BADRANE; TORDO, 2001). O filogrupo I inclui o *Rabies virus* (RABV), o *Duvenhage virus* (DUVV), o *European bat lyssavirus 1* (EBLV-1), o *European bat lyssavirus 2* (EBLV-2), o *Australian bat lyssavirus* (ABLV), o *Aravan virus* (ARAV), o *Khujand virus* (KHUV), o *Irkut virus* (IRKV) e o *Bokeloh vírus* (BBLV). O filogrupo II inclui o *Lagos bat virus* (LBV), o *Mokola virus* (MOKV) e o *Shimoni vírus* (SHIBV). O *West Caucasian bat virus* (WCBV) e o *Ikoma lyssavirus* (IKOV) são os mais divergentes entre os quatorze genótipos reconhecidos dos *Lyssavirus* e não se agrupam em nenhum dos dois filogrupos (KUZMIN et al., 2003; BOTVINKIN et al., 2003; KING et al., 2012; (MARSTON et al., 2012).

Pesquisas sugerem que *Lyssavirus* que apresentam menos que 80% de identidade na sequência de nucleotídeos para o gene N completo ou todos os 5 genes concatenados e menos 92% de similaridade na sequência de aminoácidos da nucleoproteína viral pertencem a diferentes genótipos. Mas outros critérios como topologia e consistência de árvore filogenéticas e padrões de reação com anticorpos monoclonais também devem ser utilizados como critério de classificação para os *Lyssavirus* (KISSI; TORDO; BOURHY, 1995; WUNNER, 2007; ARAI et al., 2003; KING et al., 2012).

A raiva está difundida em todos os continentes, exceto na Antártica. Os *Lyssavirus* são muito lábeis e não conseguem manter-se no ambiente. O vírus da raiva tem sido isolado da maioria das ordens de mamíferos, sendo que alguns destes atuam como reservatórios do RABV em diferentes partes do mundo, principalmente os da ordem *Carnivora* e *Chiroptera*. (RUPPRECHT; HANLON; HEMACHUDHA, 2002)

Entre os *Lyssavirus*, o RABV é o mais amplamente distribuído mundialmente e tem maior importância epidemiológica dada sua associação com um maior número de casos de encefalite por *Lyssavirus* em humanos em relação às outras espécies (RUPPRECHT et al., 2002).

Estudos moleculares e de cadeia de transmissão do RABV nas Américas têm mostrado a existência de diversos reservatórios para o genótipo 1, nos quais linhagens virais adaptadas a eles se mantêm na natureza em ciclos epidemiológicos independentes. Dentro de cada ciclo, estes diferentes reservatórios exercem um papel fundamental e específico na manutenção de cada linhagem viral. No Brasil temos reservatórios pertencentes as ordens *Carnivora*, *Chiroptera* e *Primates* (VELASCO-VILLA et al., 2002; FAVORETTO et al., 2001; CARNIELI et al., 2008; STREICKER et al., 2010; OLIVEIRA et al., 2010; PIÑERO et al., 2012; YUNG; FAVI; FERNANDES, 2012).

São admitidos dois ciclos de transmissão para a raiva, o ciclo urbano e o ciclo silvestre. O ciclo urbano tem o cão como principal reservatório e transmissor do vírus para outros cães, outros animais domésticos e para o homem. O ciclo silvestre é mantido por diferentes mamíferos silvestres e quirópteros (ACHA, 2003).

A via de transmissão do vírus é a saliva do animal infectado, que transmite o vírus através de mordeduras e/ou lambeduras. Também há casos citados de transmissão entre humanos por aerossóis em cavernas densamente povoadas por morcegos e casos de transmissão iatrogênica através de cirurgias de transplantes de órgãos (RUPPRECHT et al., 2002).

Os cães são os principais reservatórios da raiva nos países em desenvolvimento, porém, na Europa e América do Norte, onde os programas de vacinação em cães estão bem estabelecidos, o RABV mantém o seu ciclo principalmente nas espécies silvestres como raposas vermelhas, mangustos, guaxinins, gambás, chacais e morcegos (RUPPRECHT et al., 2002).

Estima-se que anualmente ocorram em torno de 55.000 óbitos humanos por raiva na Ásia e África. Na América Latina, a incidência anual da raiva por 100.000 habitantes varia entre 0 e 0,09 na América do Sul, 0 e 0,10 na América Central e 0 e 0,06 nas ilhas do Caribe.

Na grande maioria dos casos, o cão foi identificado como o animal agressor (CHILDS; REAL, 2007).

No Brasil, no período entre 1986 e 2010, ocorreram 767 óbitos por raiva humana, sendo que destes, 519 tiveram o cão como animal agressor, seguido dos quirópteros que foram responsáveis por 137 casos¹ (SVS/MINISTÉRIO DA SAÚDE).

Na América do Norte e Europa, a raiva humana é atualmente uma doença rara com poucos casos anuais (CHILDS; REAL, 2007).

Entre os quatorze genótipos do gênero *Lyssavirus*, com exceção do MOKV, todos têm os quirópteros como reservatórios (TORDO, 1996; CALISHER; ELLISON, 2012; MARSTON et al., 2012). Além disso, evidências mostram que todas as linhagens de RABV atualmente circulantes em carnívoros terrestres são originárias de linhagens específicas associadas a quirópteros (BADRANE; TORDO, 2001).

Os quirópteros são altamente móveis e a capacidade de certas espécies de se adaptar em ambiente urbano e abrigar-se em habitações humanas aumentam a probabilidade de contato com humanos e animais domésticos (UIEDA; HARMANI; SILVA, 1995). A adaptação dos morcegos insetívoros, que constituem a maior parte da população de morcegos, ao meio urbano se deve em grande parte à abundante oferta de alimento e abrigo, associada à ausência de predadores (ALMEIDA et al., 1994). O número desses animais nas áreas urbanas tem aumentado constantemente (TADDEI, 1983).

A raiva em morcegos apresenta um ciclo epidemiológico independente dos ciclos existentes nos mamíferos terrestres. A raiva em morcegos hematófagos ocorre somente na América Latina e Trindade e Tobago, onde habitam os morcegos hematófagos. A raiva em morcegos não hematófagos é registrada indistintamente nos países desenvolvidos e em desenvolvimento das Américas, representando um problema emergente de saúde pública, pela expansão das áreas de ocorrência, incluindo áreas urbanas (ACHA, 2003).

A primeira observação vinculando a raiva aos morcegos hematófagos no Brasil foi feita por Carini em 1910 (CARINI, 1911) e a primeira morte humana atribuída à mordida de morcegos vampiros foi relatada em 1931 em Trindade (PAWAN, 1959).

Episódios de raiva humana causada por morcegos hematófagos continuaram sendo relatados em muitos países da América Latina, tais como México, Peru, Venezuela e Brasil. Entre os anos de 2004 e 2005, os morcegos hematófagos foram os principais transmissores de raiva humana na América Latina, com 46 e 52 casos, respectivamente. O Brasil foi o

¹ Dados cedidos pela Secretaria de Vigilância em Saúde do Ministério da Saúde – SVS/MS.

responsável por 64 destes casos (22 em 2004 e 42 em 2005) devido ao surto de raiva humana transmitida por morcegos hematófagos ocorrido nos estados do Pará e Maranhão nestes anos. A raiva em morcegos hematófagos, além de ser um sério problema de saúde pública na América do Sul, também causa grande prejuízo econômico para a pecuária destes países (ROSA, DA et al., 2006; KOTAIT et al., 2007).

O reconhecimento dos morcegos insetívoros como reservatórios do vírus da raiva na América do Norte, ocorreu na Flórida em 1953 (SCATTERDAY; GALTON, 1954). Após este fato, inúmeros casos de raiva humana vêm sendo descritos na América do Norte, tendo como fonte de infecção morcegos insetívoros, principalmente das espécies *Lasionycteris noctivagans* e *Pipistrellus subflavus*, sendo que a maioria deste casos não tem histórico de exposição a estes animais. Casos de raiva humana nos quais foram identificadas variantes próprias de morcegos, sem evidências de mordeduras, também foram relatados em diversos outros países da Europa e do continente americano (KOTAIT et al., 2007).

No Brasil, com referência aos morcegos não hematófagos, a primeira comunicação sobre o isolamento do vírus foi realizada em 1957 no Rio de Janeiro em morcego *Phyllostomus hastatus hastatus* (SILVA, 1964). A partir desta data ocorreram outros isolamentos em diversas outras espécies de morcegos insetívoros (CUNHA et al., 2006).

No Brasil, assim como na maior parte da América Latina, apesar destes achados, a importância dos morcegos não hematófagos na epidemiologia da doença continuou pouco estudada até a década de 80, devido a presença de raiva mantida por cães e morcegos hematófagos (ACHA et al., 1985). A partir desta década, com o controle da raiva canina em muitos municípios e incorporação da tipificação molecular e antigênica aos programas de vigilância, uma apreciação da importância dos morcegos não hematófagos começou a surgir nesses países (DE MATTOS, et al., 1996, 2000; FAVORETTO et al., 2002).

Entre as 1.120 espécies de quirópteros existentes no mundo, 9 famílias, 64 gêneros e 165 espécies são encontrados no Brasil. Destas, 41 espécies, pertencentes a 25 gêneros e 3 famílias já foram diagnosticadas com o vírus da raiva (SODRÉ; DA GAMA; ALMEIDA, 2010).

Na América do Sul, pesquisas mostram que os gêneros/espécies de morcegos não hematófagos com maior importância epidemiológica para a raiva são: *Tadarida brasiliensis*, *Myotis sp*, *Lasiurus sp* e *Artibeus sp* (KOTAIT et al., 2007).

O conceito de variantes do vírus da raiva e o estudo das suas diferenças e reservatórios específicos foi consolidado com o desenvolvimento da técnica de anticorpos monoclonais (AcM) para a nucleoproteína e glicoproteína virais, no final da década de 70. Diferentes

painéis destes anticorpos permitem uma identificação e classificação rápida de isolados de *Lyssavirus* (CUNHA, 2006).

Desde então, com a utilização da caracterização antigênica para a classificação do RABV, muitos avanços foram obtidos na epidemiologia da raiva, tornando-se possível determinar a distribuição geográfica e reservatórios específicos de diferentes variantes do RABV (FAVORETTO et al., 2002).

Um painel de oito AcM dirigidos à nucleoproteína viral do RABV utilizado nas Américas, permite a classificação do RABV em 11 variantes antigênicas distintas e adaptadas a diferentes reservatórios como mostrado no quadro 1 (DIAZ et al., 1994).

Quadro 1 – Relação das variantes antigênicas do RABV encontradas nas Américas e seus respectivos reservatórios - São Paulo -2014

Hospedeiro	Variante
Cão/Mangusto	1
Cão	2
<i>Desmodus rotundus</i>	3
<i>Tadarida brasiliensis</i>	4
<i>Desmodus rotundus</i>	5
<i>Lasiurus cinereus</i>	6
Lobo do Arizona	7
Gambá Centro/Sul	8
<i>T. brasiliensis mex.</i>	9
Baja SC Gambá	10
<i>Desmodus rotundus</i>	11

Fonte: Diaz et al., 1994

No Brasil, encontram-se as variantes antigênicas 2, 3, 4 e 6, entretanto, estudos utilizando este painel de AcM também demonstram a existência de outros quatro perfis

antigênicos não compatíveis com os perfis esperados para este painel, os quais foram encontrados em diferentes gêneros de morcegos insetívoros (FAVORETTO et al., 2002; OLIVEIRA, 2009; ALBAS et al., 2011; QUEIROZ et al., 2012).

Com o desenvolvimento de técnica do sequenciamento genético para o estudo da raiva, as limitações inerentes à técnica dos AcM foram superadas, o que permitiu estabelecer uma relação definitiva entre linhagens virais intimamente relacionadas (BRASS, 1994).

O uso deste painel de AcM apresenta limitações como por exemplo, na análise de diferentes linhagens genéticas intimamente relacionadas antigenicamente além de linhagens genética não classificadas através deste painel. Como exemplo, podemos citar duas linhagens genéticas distintas brasileiras (*Myotis sp* e *T. brasiliensis*) caracterizadas antigenicamente como variante 4, sendo este um perfil antigênico relacionado erroneamente apenas a espécie *T. brasiliensis* (OLIVEIRA, 2009; ALBAS, et al, 2011; QUEIROZ et al., 2011).

Estudos genéticos recentes realizados no Chile e Argentina mostram a compartimentalização de linhagens de RABV em pelo menos 6 gêneros de morcegos insetívoros (*Tadarida brasiliensis*, *Myotis sp*, *Eptesicus sp*, *Histiotus montanus*, *Lasiurus blosseviilli* e *Lasiurus cinereus*), sendo que 3 destas linhagens (*Tadarida brasiliensis*, *Lasiurus blosseviilli* e *Lasiurus cinereus*) também são mantidas por morcegos brasileiros (PIÑERO et al., 2012; YUNG et al., 2012).

No Brasil, estudos mostram a existência de pelo menos 11 linhagens de RABV mantidas por gênero ou espécie específica de morcegos insetívoros. Pelo fato de que a maioria destas amostras serem do estado de São Paulo, pouco se conhece sobre a distribuição destas linhagens em outras regiões do país. Isto também dificulta a realização de estudos filogeográficos para a identificação de um padrão de distribuição de subgrupos dentro de uma linhagem, relacionados a distribuição geográfica de seus reservatórios (OLIVEIRA, 2009; OLIVEIRA et al., 2010; ALBAS et al., 2011; ALLENDORF et al., 2012; MENOZZI, 2012).

Estudo brasileiros evidenciaram que os morcegos dos gêneros *Myotis* e *Eptesicus* são responsáveis pela manutenção de pelo menos duas linhagens virais cada um. Além disso, gêneros ou espécies de morcegos insetívoros reservatórios de determinada linhagem podem ser encontrados infectados com linhagens específicas de outros gêneros de morcegos insetívoros, e também com a linhagem específica do morcego hematófago *D. rotundus*. Estes casos de transmissão inter espécies (spillovers) são descritos no Brasil, principalmente em morcegos do gênero *Molossus* e *Nyctinomops* (FAVORETTO et al., 2002; CUNHA, 2006; KOBAYASHI et al., 2005, 2007; OLIVEIRA, 2009; OLIVEIRA et al., 2010; ALBAS et al., 2011; QUEIROZ et al., 2011; MENOZZI, 2012; ALLENDORF et al., 2012).

Nos Estados Unidos já foram caracterizada pelo menos 18 linhagens de RABV mantidas por diferentes gêneros ou espécies de morcegos insetívoros. Como observado no Brasil, alguns gêneros ou espécies de morcegos são responsáveis pela manutenção de mais de uma linhagem viral. A transmissão inter espécies também é frequente mas de pouca importância epidemiológica para a manutenção destas linhagens, já que estes morcegos aparentemente são hospedeiros terminais para tais linhagens. Além disso, foi demonstrada a manutenção de uma mesma linhagem de RABV por três diferentes espécies de morcego insetívoro diferentes. Estas transmissões inter espécies se dão preferencialmente entre morcegos próximos filogeneticamente e de hábitos ecológico similares (STREICKER et al., 2010; STREICKER et al., 2012).

Desde o final início dos anos 2000, sabe-se que na região nordeste do Brasil, o canídeo silvestre *Cerdocyon thous* e o sagui do tufo branco *Callithrix jacchus* são reservatórios para linhagens específicas de RABV que se mantém em ciclos epidemiológicos independentes nestas espécies (FAVORETTO et al., 2002, 2013; CARNIELI et al., 2006, 2008), sendo as mesmas relacionadas a diversos casos de raiva humana ocorridos nesta região deste então. Além disso, no estado do Maranhão a linhagem do RABV específica do *C. thous* foi introduzida no meio urbano sendo mantida através da infecção de cães, que atualmente mantém esta linhagem de RABV nesta região independentemente do canídeo silvestre, além da linhagem específica do cão doméstico (KOBAYASHI, YUKI et al., 2007).

Para um melhor entendimento da epidemiologia, filodinâmica e evolução do RABV, o estudo de genomas virais completos é imprescindível. No Brasil, até o momento, apenas 6 exemplares de RABV, dois pertencentes a linhagem *D. rotundus* e um pertencente a linhagem *Cerdocyon thous*, dois pertencentes a linhagem específica de cães e um pertencente a uma das linhagens mantidas pelo gênero *Lasiurus* tiveram o seu genoma sequenciado (MOCHIZUKI et al., 2009, 2011; CAMPOS et al., 2011). Campos et al., (2011) obteve êxito ao sequenciar 11622 nucleotídeos desde a extremidade 3' UTR de um isolado de RABV pertencente a linhagem *D. rotundus*. Entre os seis genomas de RABV brasileiros sequenciados, nenhum é representativo das outras 10 linhagens circulantes em morcegos insetívoros e saguis.

O vírus da raiva (RABV) apresenta um genoma não segmentado de RNA fita-simples com polaridade negativa, fazendo com que o RNA viral não seja infeccioso por não ser capaz de ser traduzido diretamente em proteínas. O genoma completo tem 11.932 nucleotídeos (nt) no vírus fixo Pasteur vírus (PV), os quais codificam as proteínas estruturais, *N*, *P*, *M*, *G* e *L*. Estes genes apresentam-se separados por quatro regiões intergênicas não codificantes (entre o final 3' de um gene e o início 5' do próximo gene), sendo compostas de 2 nt (*N-P*), 5 nt (*P-*

M), 5 nt (P-G) e 24 nt (G-L). Estas regiões têm importante papel na regulação da expressão gênica viral (TORDO et al., 1986; FINKE; COX; CONZELMANN, 2000; KING et al., 2012).

Como em todos os vírus de RNA de sentido negativo, o genoma do RABV apresenta-se firmemente encapsidado pela nucleoproteína viral, sendo que o RNA viral é funcional para a transcrição e replicação apenas nesta forma encapsidada denominada ribonucleoproteína (RNP). O capsídio viral é formado pelo RNP acoplado ao complexo RNA polimerase viral, constituído pela RNA polimerase RNA dependente (proteína L) e pela fosfoproteína (proteína P), a subunidade não catalítica do complexo RNA polimerase (SCHNELL et al., 2010).

No envelope viral do RABV encontram-se espículas formadas por trímeros da glicoproteína viral (proteína G). Abaixo do envelope existe uma camada matriz, formada pelas proteínas M, que unem o envelope viral à ribonucleoproteína viral (RNP), sendo que a proteína M em conjunto com a RNP formam o ribonucleocapsídeo do vírus. A proteína M é a responsável pela condensação da RNP em sua forma helicoidal e também interage com a região citoplasmática da glicoproteína do RABV (MEBATSION, 2001).

Alguns vírus fixos da raiva, entre eles o Pasteur Virus (PV) e o Evelyn-Rokitnicki-Abelseth Virus (ERA) apresentam uma longa região intergênica entre os genes *G* e *L*, denominada de gene remanescente ou o pseudo gene viral (Ψ) devido ao seu considerável tamanho (423 nt) e ausência de uma janela de leitura (ORF) detectável, (TORDO et al., 1986; TORDO, N et al., 1986; TORDO, 1996; WUNNER, 2007), sendo a região é a mais variável no genoma destes vírus (SACRAMENTO; BOURHY; TORDO, 1991). Isto ocorre devido ao fatos destes vírus possuírem uma cauda poli A interna no RNA mensageiro do gene *G* em relação cauda poli A encontrada na maioria absoluta deste gene nos RABV, sendo que neste caso, serão transcritos dois tipos de RNA mensageiro a partir do gene *N*, um curto que termina na cauda poli A interna e outro mais longo, igualmente ao restante dos RABV (WUNNER, 2007).

Na porção 3' do genoma encontra-se uma sequência de 55 a 58 nucleotídeos não codificantes, conhecida como *leader* (Le). Logo em seguida encontram-se as sequências que codificam para as cinco proteínas virais. No final do genoma, existe uma sequência de 70 nucleotídeos na região 5' denominada *trailer* (Tr). (TORDO, NOEL et al., 1986; TORDO, N et al., 1986). As sequências Le e Tr apresentam papel chave nos processos de transcrição e replicação viral (SCHNELL et al., 2010).

A proteína N (nucleoproteína viral) contém 450 aminoácidos possuindo um peso molecular de aproximadamente 57.000 daltons, sendo a principal proteína viral e a principal

proteína do nucleocapsídeo. Em termos de similaridade de sequência de aminoácidos, a proteína N é a mais conservada entre os sete genótipos dos *Lyssavirus*, apesar de um grau relativamente alto de diversidade genética ser encontrada em algumas pequenas regiões do gene N nos diferentes genótipos. Por essa razão, para a detecção do vírus da raiva por RT-PCR, tem sido mais utilizado o gene da nucleoproteína viral (WUNNER, 2007).

Uma importante razão para a sua maior conservação de aminoácidos, principalmente em regiões específicas, são as funções chaves para a replicação viral exercidas por essas regiões. Entretanto, diferenças de aminoácidos na nucleoproteína fornecem epítomos específicos e únicos capazes de diferenciar os *Lyssavirus* em seus diferentes genótipos através dos padrões de reação de anticorpos monoclonais (AcM) para este epítomos (WUNNER, 2007).

A proteína N possui quatro sítios antigênicos (I a IV), localizados nos segmentos de aminoácidos de posições 358-367 (sítio I), 373-383 (sítio II), 313-337 (sítio III), 359-366 e 375-383 (sítio IV) (TORDO, 1996; WUNNER, 2007).

A proteína N representa o principal alvo para os linfócitos T auxiliares que tem reação cruzada entre os *Lyssavirus*. Vários destes epítomos foram mapeados e testados quanto à capacidade de estimular linfócitos T auxiliares específicos para o RABV in vivo. Um destes epítomos, denominado 31D, que corresponde aos resíduos 404-418 na nucleoproteína do RABV, se trata de um epítomo imunodominante capaz de estimular a produção de linfócito T auxiliares anti-RABV in vitro (WUNNER, 2007).

A proteína P do RABV contém 297 aminoácidos (891 nucleotídeos) na amostra PV sendo bastante conservada (>97%) entre os RABV. A proteína P (fosfoproteína) dos *Lyssavirus*, assim como em outros vírus de genoma de sentido negativo, se apresenta em diferentes formas fosforiladas, sendo uma proteína multifuncional, interagindo com a nucleoproteína (N-P) e atuando como componente chave para a replicação do genoma viral, quando associada a RNA-polimerase RNA dependente (proteína L) do RABV (P-L). A interação N-P e P-L acontece em domínios proteicos específicos e independentes da fosfoproteína viral (WUNNER, 2007).

A proteína P atua como chaperone para a proteína N recém sintetizada, prevenindo sua polimerização e ligações não específicas à moléculas de RNA celular, além de dirigir a encapsidação do RNA viral pela nucleoproteína viral. Como subunidade do complexo RNA polimerase (P-L), a proteína P atua como cofator não catalítico na transcrição e replicação do genoma viral, estabilizando a polimerase viral e dirigindo o acoplamento da polimerase ao molde de RNA (WUNNER, 2007).

A proteína P também interage com proteínas celulares que influenciariam ou ajudariam a regular o tropismo celular e a movimentação célula a célula do RABV, ou ainda inibir a resposta imune inata, que geralmente interromperia a replicação viral. Uma destas proteínas é a cytoplasmic dynein light chain (LC8), a qual está relacionada ao transporte intracelular de organelas. O sítio de interação da fosfoproteína do RABV com a LC8 está localizado em sua porção amino terminal, entre os nucleotídeos 138 a 172 (WUNNER, 2007; SCHNELL et al., 2010).

A proteína P do RABV também é responsável por inibir a ativação do fator de regulação de interferon 3 (IRF-3), molécula responsável pela inicialização da resposta por interferons (IFN), através da prevenção da fosforilação de vários resíduos de serina na região carboxi terminal do IRF-3, necessária para a formação dos dímeros desta proteína, necessários para a ativação da resposta por interferons. Desta maneira, a proteína P se mostra necessária e suficiente para prevenir a resposta por interferons nas células infectadas (WUNNER, 2007; SCHNELL et al., 2010).

A proteína M do RABV e outros genótipos dos *Lyssavirus* é a menor das proteínas do vírion, possuindo no RABV 202 aminoácidos (25 kDa) (WUNNER, 2007; SCHNELL et al., 2010).

A proteína M forma dímeros (54kDa; 10-20% do total de M na célula e 20 a 30% no vírion) os quais estão fortemente associados com a proteína G. A proteína M liga-se e condensa o núcleo do nucleocapsídeo nascente em complexo proteico helicoidal RNP-M, formando um revestimento em torno deste e conferindo ao RABV seu formato de bala de revólver. Entre 1.200 a 1.500 moléculas de M se ligam ao RNP viral (WUNNER, 2007; SCHNELL et al., 2010).

A proteína M também é responsável pela ligação do RNP viral a membrana celular da célula hospedeira, na periferia do citoplasma celular, onde inicia o processo de brotamento do RABV através da membrana plasmática. Um motivo proteico rico em prolina (PPPY ou PY) localizado entre os resíduos 35 a 38 da proteína M parece estar associado ao processo de brotamento do RABV. Estes quatro aminoácidos estão localizados dentro da região compreendida nos primeiros 14 de região N-terminal da proteína M, a qual se apresenta bastante conservada entre os RABV (WUNNER, 2007).

A glicoproteína G do vírus da raiva é uma proteína de membrana com 505 aminoácidos, traduzidas de um mRNA que codifica 524 aminoácidos, sendo a proteína de fusão, que media a entrada do vírus na célula. Além disso, é fundamental para a resposta

imune contra o vírus da raiva, sendo responsável pela indução de anticorpos neutralizantes, sendo alvo destes e dos linfócitos T *helper* e citotóxicos (WUNNER, 2007).

A glicoproteína G tem importante papel na adsorção viral a receptores específicos nas células dos hospedeiros, na indução de anticorpos neutralizantes contra o RABV e na patogênese e patogenicidade da doença, sendo composta por quatro regiões distintas: a região peptídeo sinal (PS), o endodomínio (ED), o domínio transmembrânico (TM) e o domínio citoplasmático (DC) (WUNNER, 2007).

Os primeiros 19 aminoácidos da região N-terminal da glicoproteína do RABV contemplam o PS, o qual é responsável pelo transporte da proteína G nascente dentro do complexo de golgi rugoso até a membrana plasmática, antes da sua clivagem dentro desta organela. A proteína G transmembrânica resultante é organizada em trímeros (três monômeros com 65 kDA cada) dentro do complexo de golgi, onde mais tarde formarão as espículas triméricas incorporadas a membrana plasmática e a superfície do vírion. As espículas de G no envelope viral se estendem 8,6 nm a partir do envelope viral e representam a maior superfície proteica no vírion. (WUNNER, 2007).

Cada proteína G está ancorada ao envelope viral por uma região de 22 aminoácidos (na verdade são 23), o domínio transmembrânico (TM), localizado entre os resíduos 439 e 461. A porção C-terminal (últimos 44 aminoácidos) da proteína G, o domínio citoplasmático (CD), se estende dentro do citoplasma plasmático a partir da membrana celular da célula infectada onde interage com a proteína M ligada os RNP, para completar o processo de montagem do vírion (WUNNER, 2007).

O ectodomínio (ED) da proteína G (resíduos 1 a 439 da glicoproteína madura), a porção que se estende a partir da membrana plasmática e superfície da partícula viral madura é a principal região da molécula por uma variedade de interações funcionais (WUNNER, 2007).

As espículas triméricas da glicoproteína são responsáveis por interações com os sítios de ligação nas membranas celulares (receptores), e portanto, com importância na patogênese viral pela ligação as células apropriadas para a infecção. Elas também são responsáveis pela fusão induzida por baixo pH do envelope viral com as membranas plasmática e endossomal da célula hospedeira nas fases iniciais do ciclo de vida do RABV. Além disso, ela é crítica para a indução da resposta imune humoral do hospedeiro à infecção do RABV como alvo de anticorpos neutralizantes, bem como para a resposta por linfócitos T auxiliares e citotóxicos específicos para o RABV (WUNNER, 2007).

A virulência de um fenótipo do RABV está correlatada com um único aminoácido específico na posição 333 da glicoproteína do RABV. A arginina (R) 333 ou lisina (K) 333 de vírus selvagens determinam a virulência ou padrões neuroinvasividade do RABV no CNS. Variantes dos vírus com as substituições glutamina (Q), isoleucina (I), glicina (G), metionina (M) ou serina (S) nesta posição expressam fenótipos menos patogênicos ou avirulentos comparados com os vírus com R ou K na posição 333, em camundongos adultos imunocompetentes. Embora estas mutações reduzam ou até mesmo anulem a neuroinvasividade do RABV, elas não influenciam na habilidade do vírus se multiplicar em cultura de células, porém pode afetar a velocidade de propagação do vírus célula a célula, e a capacidade de infectar neurônios motores *in vivo* e *in vitro*. Esta única substituição de aminoácido pode modificar o espectro de hospedeiros de determinado vírus determinando a escolha de vias neuronais que o vírus utiliza para alcançar o CNS, bem como a distribuição do vírus em diferentes regiões do cérebro (WUNNER, 2007).

Oito sítios antigênicos foram identificados no domínio externo da proteína G (I-VI, “a” e G1). Os sítios I, III, VI e “a” envolvem os aminoácidos nas posições 231, 330-338, 264 e 342-343, respectivamente. O sítio II é descontínuo, localizado nas posições 34-42 e 198-200 ligadas por pontes dissulfeto (TORDO, 1996).

A proteína L do RABV contém 2.142 aminoácidos na amostra PV (2.127 na amostra SAD-B19 e MOKV), possuindo 244 kDa, sendo codificada pelo quinto gene do genoma do RABV, compreendendo mais da metade do seu potencial codificante (54%). Esta proteína se apresenta como o componente catalítico da RNA-polimerase viral e junto com o cofator não catalítico P, é responsável pela maioria das atividades enzimáticas envolvidas na transcrição e replicação do RNA viral (SCHNELL et al., 2010).

Como em todos os vírus de RNA sentido negativo, a RNA polimerase associada ao vírion executa um papel crítico na infecção viral, iniciando a transcrição primária do genoma RNA, após a liberação do RNP no citoplasma da célula hospedeira. Os processos enzimáticos da transcrição incluem a iniciação e alongação da região Leader + (Le +), transcrição de RNAs mensageiros, bem como as suas modificações pós-transcricionais tais como 5' capping, metilação e 3' poliadenilação (SCHNELL et al., 2010).

A sequência de eventos no ciclo de vida do RABV (replicação *in vitro* e *in vivo*) pode ser dividido em três fases. A primeira, que inclui a adsorção do vírion a receptores celulares das células do hospedeiro, sua entrada via fusão direta externamente com a membrana plasmática e internamente com membranas endossomais celulares, seguidos do seu desnudamento e liberação do RNP helicoidal no citoplasma celular (SCHNELL et al., 2010).

A segunda fase incluem a transcrição e replicação do genoma viral e síntese de proteínas, seguida da terceira fase que incluem a montagem viral e seu egresso da célula infectada (SCHNELL et al., 2010).

O ciclo de vida do RABV se inicia com sua adsorção a superfície de uma célula alvo, seguido de sua penetração nesta célula por endocitose seguido da fusão do envelope viral ao envelope endossomal para a liberação do RNP no citoplasma celular. Muito provavelmente o vírus liga-se a moléculas receptoras celulares que possibilitam sua entrada em células nervosas a partir da junção neuromuscular, de onde são transportadas até o axônio do neurônio infectado onde completam o seu ciclo de vida (SCHNELL et al., 2010).

O primeiro passo do ciclo de vida do RABV é a sua adsorção a receptores específicos da célula hospedeira, sendo que este mecanismo é mediado pela glicoproteína viral, porém quais seriam estes receptores ainda não está totalmente esclarecido (SCHNELL et al., 2010).

O receptor nicotínico de acetilcolina (nAChR) foi o primeiro potencial receptor identificado para o RABV, mas devido ao fato de que os nAChRs só são encontrado nas membranas pós sinápticas das junções neuromusculares é pouco provável que este receptor seja utilizado para a entrada inicial do RABV nas células hospedeiras. Porém, existe a hipótese de que os nAChRs sejam responsáveis por auxiliar na infecção de neurônios adjacentes e na otimização da infecção inicial de células musculares estriadas pelo RABV (SCHNELL et al., 2010).

Outros potenciais candidatos a receptor para o RABV são a molécula de adesão da célula neuronal (NCAM) e o receptor de baixa afinidade para o fator de crescimento nervoso (p75NTR). O NCAM não tem papel essencial na adsorção viral, porém auxilia no processo, enquanto o p75NTR parece necessitar de outras moléculas para exercer este papel. Carboidratos, lipídios e gangliosídeos tem também sido sugeridos como receptores para o RABV, mas suas interações com o RABV não são bem compreendidas (SCHNELL et al., 2010).

Nenhuns dos potenciais receptores para o RABV parecem ser essenciais *in vitro*, sendo que o fato de uma única substituição na glicoproteína viral pode fazer com que o vírus perca a sua neuroinvasividade e reduzir sua velocidade de adsorção, só poderia ser explicado pela utilização de dois receptores diferentes. Diferenças na velocidade de adsorção e espalhamento em culturas celulares de vírus de morcegos e vírus vacinais também sugerem que diferentes receptores podem ser usados (SCHNELL et al., 2010).

Qualquer que seja o receptor celular utilizado pelo RABV para a sua adsorção e fusão com a célula hospedeira, é evidente que o vírus não está limitado a apenas uma alternativa

para completar o primeiro estágio do seu ciclo de vida no animal infectado (WUNNER, 2007).

Depois da sua ligação ao receptor celular, a internalização do vírus se dá por endocitose mediada por receptores, sendo que a proteína G media o mecanismo de fusão do envelope viral com a membrana endossomal e liberação do RNP viral a partir das vesículas endossomais. O mecanismo de fusão do envelope viral com a membrana endossomal é dependente de baixo pH (6.3) e envolve uma série de mudanças conformacionais específicas na glicoproteína viral (SCHNELL et al., 2010).

O sítio de entrada do RABV nos axônios não fornece o ambiente bioquímico ideal para a sua síntese proteica e dessa maneira o vírus necessita alcançar o corpo do neurônio para realizar sua transcrição e replicação. Dois diferentes mecanismos têm sido propostos para o transporte do RABV através do axônio, o transporte apenas do capsídeo viral ou do vírion completo (SCHNELL et al., 2010).

O transporte apenas do capsídeo viral dependeria de sua liberação no citoplasma celular logo após a sua internalização além de interações específicas do RNP com a maquinaria de transporte celular. O reconhecimento da interação entre a proteína P e a molécula dineína de cadeia leve 8 (LC8) levou a hipótese de que o vírus usa o complexo motor citoplasmático da dineína para o seu transporte intracelular. Entretanto, a deleção na fosfoproteína do sítio de ligação para a LC8 não afeta o transporte viral até o CNS, porém a LC8 interfere direta ou indiretamente na transcrição primária do RABV mas não no seu transporte (SCHNELL et al., 2010).

Já foi demonstrado o transporte da partícula viral inteira em endossomos através do axônio, embora o provável mecanismo envolvido neste transporte não foi elucidado. Este mecanismo dependeria de interações da glicoproteína com proteínas celulares, mas o fato da proteína G estar localizada no interior do endossomo impediria a sua interação direta com algum mecanismo de transporte celular. Baseado nesta observação especulou-se que a entrada do RABV no endossomo, a qual é mediada pela glicoproteína do RABV determinaria a direção e a força motora deste mecanismo de transporte. Isto sugere que a natureza do endossomo formado na internalização do RABV determinaria o mecanismo por trás do transporte do RABV nas células nervosas (SCHNELL et al., 2010).

A expressão gênica e replicação do RABV é altamente regulada e difere em alguns aspectos em relação a outros rabidovírus, inclusive o vírus da estomatite vesicular (VSV), o vírus protótipo para o estudo da família *Rhabdoviridae*. Embora pareça que o VSV produza uma maior quantidade de vírus no menor tempo possível, o RABV procura preservar a célula

infectada por dois motivos. Primeiramente, a indução de cito toxicidade no neurônio infectado provavelmente dificultaria o transporte neuronal do RABV, além disso, a produção de altos níveis de proteínas virais, especialmente a glicoproteína, poderia induzir uma forte resposta do sistema imune humoral, o que provavelmente neutralizaria o vírus antes que ele pudesse completar o seu ciclo de vida (SCHNELL et al., 2010).

Portanto, o RABV possui mecanismos regulatórios para otimizar a produção de seus componentes virais em níveis baixos o suficiente para não serem reconhecidos precocemente pelo sistema imune ou interferir nas funções vitais da célula hospedeira (SCHNELL et al., 2010).

O genoma RNA não segmentado fita simples dos lissavírus não pode ser traduzido diretamente em proteínas. Por isso, o primeiro evento no processo da replicação do RABV é a sua transcrição para produzir os RNAs mensageiros monocistrônicos das cinco proteínas do RABV (WUNNER, 2007).

Na segunda fase do ciclo de vida do RABV, a transcrição do genoma viral de RNA inicia-se no citoplasma do neurônio infectado após a liberação do RNP viral das vesículas endossomais. Sua estrutura antes fortemente enrolada apresenta-se agora em um formato helicoidal fracamente enrolado, o que assegura a facilitação da transcrição e replicação do RNA viral. Pesquisas sugerem que a replicação e transcrição do RABV ocorram nos corpúsculos de Negri, que são inclusões citoplasmáticas patognomônicas para a raíva, encontradas em neurônios infectados e tem sido a muitos anos utilizados para a realização do diagnóstico desta doença (SCHNELL et al., 2010).

O processo de transcrição viral no RABV é realizado através da interação do RNA viral complexado com a proteína N (RNA-N) com o complexo RNA polimerase RNA dependente (proteína L e cofator proteína P). A polimerase viral associada ao vírus inicia a transcrição na extremidade 3' do genoma ou em outros sítios de iniciação em RNAs mensageiros a jusante no genoma. Este processo de replicação é denominado transcrição primária uma vez que utiliza o RNA liberado no citoplasma celular e não requer proteínas do hospedeiro, nem a síntese prévia de alguma proteína viral (WUNNER, 2007; SCHNELL et al., 2010).

Cada um dos cinco genes do genoma produzem em ordem sequencial transcritos monocistrônicos de RNAm que oportunamente serão traduzidos em cada uma das cinco proteínas virais. A cada região intergênica, a polimerase interrompe a transcrição e estima-se que neste momento, entre 20 a 30% dos complexo RNA-polimerase se dissociam do nucleocapsídeo, fazendo com que cada vez menos moléculas de L permaneçam acopladas ao

molde RNA-N após cada região intergênica para continuar a transcrição. Consequentemente, o número de RNAs mensageiros sintetizados para os genes a jusante no genoma vai sendo gradualmente reduzido à medida que as moléculas de L se dissociam do RNA-N, num processo de auto-regulação da expressão gênica. O número de nucleotídeos das regiões intergênicas (2, 5, 5 e 24 nucleotídeos) também regulam os níveis de transcrição dos RNAs virais (WUNNER, 2007; SCHNELL et al., 2010).

Os RNAs mensageiros monocistrônicos são então traduzidos nas proteínas virais, num processo que envolve os ribossomos celulares, fatores de iniciação e elongação e RNA transportadores do hospedeiro. A organização e características do genoma do RABV são críticas nos processos de transcrição e replicação do RABV (SCHNELL et al., 2010).

Quatro dos cinco RNAs mensageiros virais, N, P, M e L são traduzidos em ribossomos livres no citoplasma enquanto o G-RNA é traduzido em ribossomos ligados ao retículo endoplasmático rugoso. A proteína G nascente é introduzida ao ser traduzida dentro do lúmen do retículo endoplasmático rugoso (RER) onde ocorrem os eventos finais de sua montagem (WUNNER, 2007; SCHNELL et al., 2010).

Enquanto estão dentro dos RER, os monômeros de G sofrem modificações em determinados resíduos de asparagina através de glicosilações e processamento de N-glicanos necessários para a formação dos trímeros funcionais da glicoproteína que serão expostos na superfície da membrana das células hospedeiras. A formação dos trímeros de glicoproteína se dá no complexo de golgi pela adição de monossacarídeos por glicosidases e glicosiltransferases (WUNNER, 2007; SCHNELL et al., 2010).

Uma vez que existam concentrações ideais de RNA viral e proteínas N, P e L acumuladas no interior da célula infectada, tem início a formação do complexo RNP viral e montagem das partículas virais, processos que continuarão enquanto a célula permanecer metabolicamente ativa (WUNNER, 2007; SCHNELL et al., 2010).

As últimas etapas do ciclo de vida do RABV são a montagem dos componentes virais e a liberação ou brotamento dos vírions juntamente com o envelope lipoprotéico oriundo da célula infectada. Para que este último fenômeno ocorra, é necessário que o nucleocapsídeo viral seja envolvido pela membrana celular da célula hospedeira, mas o mecanismo por trás do transporte do nucleocapsídeo até o local de brotamento ainda é desconhecido. Em contraste com alguns outros vírus envelopados, como o HIV, onde as proteínas do capsídeo viral são responsáveis pelo processo de brotamento das partículas virais, no RABV são as proteínas M e G que exercem importante papel nesta função (SCHNELL et al., 2010).

A montagem das partículas virais começa com a encapsidação de RNA viral pelo complexo proteico N-P, para a formação do complexo RNA-N-P. Quando N está ligada ao fosfato da ribose o RNA viral torna-se totalmente protegido da degradação por ribonucleases celulares e possivelmente da resposta imune inata contra RNAs virais. Apesar de todo o conhecimento dos processos envolvidos na montagem viral, pouco se sabe sobre o mecanismo pelo qual a proteína L é complexada para terminar a formação do RNP (RNA-N-P-L). Acredita-se que como em outros vírus da ordem *Mononegavirales*, a proteína P seria a responsável por mediar a ligação da proteína L ao complexo RNA-N-P (WUNNER, 2007).

A proteína M é uma proteína multifuncional envolvida em vários processos relacionados a formação da progênie viral, sendo o próximo componente a ser complexado ao RNP viral (RNA-N-P-L). Primeiramente a proteína M coordena a mudança no balanço dos processos de transcrição e replicação viral realizados pelo RNP viral, com aumento da replicação viral e diminuição da transcrição. Esta regulação pode estar relacionada a condensação helicoidal do RNP viral direcionada pela proteína M, fazendo com que a atividade da polimerase viral seja interrompida. Após a realização deste processo, a proteína M irá transportar o RNP viral condensado até locais específicos da membrana celular onde estão transpassados os trómeros da glicoproteína viral, onde ocorrerão as interações entre a proteína M e o domínio citoplasmático da proteína G as quais são essenciais para estabilizar os trómeros da glicoproteína na superfície do vírion para que o brotamento da partícula viral a partir da célula hospedeira infectada seja eficiente (WUNNER, 2007).

A replicação do genoma do RABV apresenta fidelidade limitada, uma vez que a RNA polimerase viral não tem atividade corretiva de inserção dos nucleotídeos, o que leva a produção de populações de genomas virais distintos que compartilham uma origem comum. Estas mutações ocorrem em diferentes taxas, variando entre 10^{-4} a 10^{-5} substituições por ciclo, dependendo da região do genoma viral. Vários outros fatores podem estar envolvidos na geração de heterogeneidade das sequências de RNA do RABV, como por exemplo, duração da infecção, via de transmissão, carga viral, resposta imunológica do hospedeiro e interações com proteínas virais (KISSI et al., 1999).

Considerando uma distribuição randômica das mutações ao longo do genoma, os diferentes genomas formam populações de mutantes, as quais aumentam rapidamente com os sucessivos ciclos de infecção em outros hospedeiros. Como resultado desta alta instabilidade genômica, estas populações de mutantes podem apresentar mutações específicas que podem caracterizá-los como genótipos distintos que podem ser específicos para determinados reservatórios. Por exemplo, várias mutações identificadas em regiões codificantes do genoma

estão correlacionadas com o tropismo seletivo por neurônios, ou mesmo pela ausência de patogenicidade em determinados hospedeiros apresentados pelo RABV (WUNNER, 2007).

A mutação é a principal fonte de variabilidade genética, onde a seleção natural, a deriva genética, o fluxo genético e a recombinação atuam moldando a estrutura genética em todas as populações. Embora exista uma vasta literatura detalhando como os padrões e processos de mutação diferem entre as espécies, os determinantes bioquímicos e evolucionários responsáveis pelas taxas em que ocorrem estas mutações continuam a ser mal compreendida para a maioria dos organismos (DUFFY; SHACKELTON; HOLMES, 2008).

Talvez, o fato mais importante para o entendimento da evolução dos organismos seja como as taxas de mutação geradas nos genomas dos organismos relacionam-se ao ritmo de mudança evolutiva no nível das populações, as quais estão refletidas nas taxas de substituição de nucleotídeos ao longo do tempo. Quando estas mutações são neutras ou quase neutras, há uma relação simples entre a taxa em que elas são geradas em um genoma (a taxa de mutação) e então fixadas ao nível de população (a taxa de substituição). Os desvios da esperada neutralidade podem revelar aspectos fundamentais do processo evolutivo, incluindo a ação da seleção natural e variações nas dinâmicas de replicação (DUFFY et al., 2008).

Muitos dos dados que se conhece sobre as taxas de mutação e substituições são relativos à história de genes isolados, sendo assim de pouca utilidade para a compreensão da história evolutiva do organismo, particularmente quando as taxas de recombinação são altas. Consequentemente, uma perspectiva genômica das taxas de mutação e substituição é fundamental para um total entendimento destas dinâmicas evolutivas (DUFFY et al., 2008).

As mutações são imprescindíveis para a evolução dos organismos através da seleção natural, na medida em que a variabilidade genética entre indivíduos pode levar ao surgimento de diferentes fenótipos dentro de uma população, os quais podem ou não se mostrar mais adaptados ao seu habitat. Em casos de mudanças inevitáveis de ambientes, é interessante analisar se os efeitos benéficos de mutações são específicos para o habitat (especialização) em questão ou se tais mutações conferem vantagens adaptativas em outros tipos de ambientes, incluindo aquele ainda não ocupados (generalização). Os organismos especialistas são aqueles capazes de prosperar em condições ecológicas específicas, enquanto os organismos generalistas podem ocupar uma grande variedade de nichos ecológicos (WASIK; TURNER, 2013).

Muitas das características de vírus de rápida evolução, aqueles com o genoma baseado em fitas simples de DNA ou RNA, os tornam sistemas ideais para o estudo das variações das

taxa de evolução, sendo que o grande conhecimento sobre a biologia molecular, replicação e ciclos de vida de vírus específicos tornam as inferências relativas à integração dos processos evolutivos, fisiológicos e estruturais. (DUFFY et al., 2008).

Embora não existam registros fósseis de vírus, sua evolução pode ser frequentemente observada a partir da escala de tempo humana, portanto eles podem ser tratados como “populações evolutivamente mensuráveis”. Tais mensurações são realizadas através de análises de coalescência em amostras seriadas pelo tempo para estimar as taxas de substituição. Além disso, são conhecidas as taxa de mutação e substituição de muitos vírus, o que torna possível a realização de estudos comparativos (DUFFY et al., 2008).

Assim sendo, a caracterização molecular através do sequenciamento de DNA do genoma completo de linhagens gêneros específicas de vírus da Raiva mantidas por diferentes reservatórios brasileiros é fundamental para a confirmação da presença dos múltiplos ciclos endêmicos, estudar o potencial transmissão inter espécies e realização de estudos evolutivos e filodinâmicos. A coexistência de praticamente todos os reservatórios brasileiros de RABV com populações humanas e torna imprescindível a compreensão da epidemiologia molecular da raiva nestas regiões.

2 OBJETIVOS

Face à necessidade de se descreverem os genomas completos de diferentes linhagens de RABV circulantes em diferentes reservatórios do RABV existentes no Brasil este estudo teve por objetivos:

- Padronizar a técnica de RT-PCR para a amplificação do genoma completo do vírus da raiva.
- Estudar a existência de heterotaquia entre os genomas dos RABV mantidos no ciclo aéreo e terrestre da raiva.
- Determinar o gene do RABV de maior acurácia para a realização de estudos filogenéticos.
- Determinar com base na distância patrística o padrão de variabilidades dos cinco genes do RABV em conjunto e separadamente nos ciclos aéreo e terrestre do RABV.
- Estudar o regime de seleção e a taxa de substituição de nucleotídeos por sítios atuantes cada um dos 5 genes estudados, bem como para o pseudo gene.
- Estudar número efetivo de códons utilizados, o índice de adaptação de códons e o uso relativo de códons sinônimos para os cinco genes do RABV nos ciclos aéreo e terrestre.
- Estimar o tempo de ancestral comum mais recente entre os RABV mantidos ciclo aéreo e o ciclo terrestre.

3 MATERIAL E MÉTODOS

Para a realização desta pesquisa foram utilizados os seguintes materiais e métodos:

3.1 AMOSTRAS

Foram utilizadas 22 amostras de sistema nervoso central (SNC), de primeira passagem em camundongos inoculados com as seguintes linhagens brasileiras de RABV, classificadas de acordo com Oliveira et al. (2010); Almeida et al. (2011) e Carnieli et al. (2008) nas seguintes linhagens genéticas: *Myotis* I, *Eptesicus* I, *Eptesicus* II, *Nyctinomops*, *Tadarida brasiliensis* South America, *D. rotundus*, *Cerdocyon thous* e *Callithrix jacchus*.

A amostra fixa CVS-11 foi utilizada como controle positivo das reações realizadas bem como para a padronização das técnicas de RT-PCR para a amplificação do genoma do RABV (Quadro 2).

Quadro 2 – Relação das amostras utilizadas nesta pesquisa - São Paulo 2014

Amostra	Linhagem	Hospedeiro	Município
1770/2012	<i>Callithrix jacchus</i>	<i>Callithrix jacchus</i>	Eusébio-CE
5402/2007	<i>Callithrix jacchus</i>	<i>Callithrix jacchus</i>	Recife-PE
6634/2008	<i>Callithrix jacchus</i>	<i>Callithrix jacchus</i>	São Benedito-CE
742/2013	<i>Cerdocyon thous</i>	<i>Homo sapiens</i>	São José do Ribamar-MA
7841/2009	<i>Cerdocyon thous</i>	<i>Canis familiaris</i>	Itapetim-PE
2990/2013	<i>Desmodus rotundus</i>	<i>Desmodus rotundus</i>	Belo Horizonte-MG
2991/2013	<i>Desmodus rotundus</i>	<i>Desmodus rotundus</i>	Belo Horizonte-MG
2992/2013	<i>Desmodus rotundus</i>	<i>Desmodus rotundus</i>	Belo Horizonte-MG
4005/2010	<i>Desmodus rotundus</i>	<i>Bos taurus</i>	Piracaia-SP
230/2010	<i>Eptesicus I</i>	<i>Eptesicus furinalis</i>	Valinhos-SP
346/2010	<i>Eptesicus I</i>	<i>Eptesicus furinalis</i>	Tambaú-SP
512/2009	<i>Eptesicus I</i>	<i>Eptesicus furinalis</i>	Ribeirão Preto-SP
3208/2006	<i>Eptesicus II</i>	<i>Eptesicus furinalis</i>	Vinhedo-SP
163/2010	<i>Myotis</i>	<i>Myotis nigricans</i>	Caieiras-SP
497/2009	<i>Myotis</i>	<i>Myotis nigricans</i>	Campinas-SP
1400/2010	<i>Myotis</i>	<i>Myotis nigricans</i>	Campinas-SP
412/2010	<i>Nyctinomops</i>	<i>Nyctinomops laticaudaus</i>	Barretos-SP
542/2010	<i>Nyctinomops</i>	<i>Nyctinomops laticaudaus</i>	Ribeiro Preto-SP
4431/2010	<i>Nyctinomops</i>	<i>Tadarida brasiliensis</i>	Nova Independência-SP
350/2010	<i>Nyctinomops</i>	<i>Nyctinomops laticaudaus</i>	Conchal-SP
1586/2010	<i>Tadarida brasiliensis</i>	<i>Tadarida brasiliensis</i>	São Bernardo do Campo-SP
3176/2009	<i>Tadarida brasiliensis</i>	<i>Tadarida brasiliensis</i>	Santo André-SP
CVS-2011	Fixa	<i>Mus musculus</i>	São Paulo

Com exceção da amostra fixa CVS-11, as 21 amostras restantes são representativas de 8 linhagens brasileiras de RABV mantidas por 7 diferentes reservatórios, provenientes de cinco estados brasileiros. As linhagens utilizadas neste estudo foram classificadas através do sequenciamento parcial dos genes N e/ou G (FAVORETTO et al., 2001; CARNIELI et al., 2008; OLIVEIRA et al., 2010; ALMEIDA et al., 2011), sendo os isolados armazenados a -80°C até a sua utilização.

As 21 amostras de campo utilizadas neste estudo foram enviadas ao Laboratório de Diagnóstico da Raiva do Instituto Pasteur no período de 2006 a 2013. (Quadro 2).

A classificação em gênero/espécie dos morcegos estudados foi realizada através das chaves taxonômicas elaboradas por Vizotto e Taddei (1973) e Gregorin e Taddei (2002). Os hospedeiros dos vírus provenientes de canídeos e primatas foram classificados em seus locais de origem.

3.2 SÍNTESE DE DNA COMPLEMENTAR (cDNA) DE TAMANHO GENÔMICO

As 22 amostras de SNC de camundongos inoculados com as linhagens de RABV estudadas nesta pesquisa foram submetidas às técnicas de extração de RNA, transcrição reversa e polimerização em cadeia pela polimerase (RT-PCR) visando a amplificação do c-DNA anti genômico do RABV.

Como controle negativo foi utilizada desde a fase de extração do RNA até a amplificação do cDNA água ultra-pura livre de DNase e RNase.

3.2.1 Extração de RNA

Para a extração de RNA total foram utilizados 0,6 g do SNC de camundongos inoculados com os RABV descritos no quadro 2. Como controle negativo foram utilizados 300µl de água ultra pura livre de DNase e RNase.

A extração de RNA total do SNC das amostras estudadas, bem como dos controles positivo e negativo, foi realizada com o método do TRIzol (Invitrogen™) seguindo-se as instruções do fabricante e resuspendidas em 50 µL de água livre de DNase/RNase.

Após sua extração, o RNA total foi acondicionado em freezers -20 °C até o momento de sua utilização.

3.3 TRANSCRIÇÃO REVERSA (RT) PARA A OBTENÇÃO DO DNA COMPLEMENTAR (cDNA) AO ANTI-GENOMA COMPLETO DO RABV

Para a realização desta técnica foram utilizados RNAs totais extraídos de SNC de camundongos inoculados com os isolados estudados nesta pesquisa, obtidos como descrito no item 3.3. Para cada lote de Transcrição Reversa realizado, foi utilizado como controle negativo a água ultra pura livre de DNase e RNase utilizada na etapa de extração de RNA do mesmo lote.

O protocolo de transcrição reversa para a obtenção do c-DNA complementar ao genoma completo do RABV foi padronizado utilizando como referência a reação padronizada por Oliveira (2008). A Transcrição reversa foi realizada com apenas o *primer* anti-senso “Final”, descrito por Campos et al. (2011) e dirigido aos primeiros 21 nucleotídeos da extremidade 3’ UTR (Trailer) do antigenoma do RABV (Quadro 3).

Quadro 3 – Primers utilizados para a realização das reações de Transcrição Reversa do genoma completo do RABV, PCR controle, PCR do genoma completo do RABV e sequenciamento de DNA de suas extremidades 3’ e 5’ UTR.-São Paulo – 2014

Primers	Sentido	Sequência	Gene/região	Posição
Início 1-20	senso	5’ ACGCTTAACAACAARATCARAG 3’	Leader	1-22
Nes_c	anti-senso	5’ GCWATCAGGATTCCATAGCT 3’	N	341-360
304	anti-senso	5’ TTGACGAAGATCTTGCTCAT 3’	N/P	1514-1533
VT1551	senso	5’ GTCGTATCTTCATGGGAG 3’	G	4842-4859
LA3	Anti-senso	5’ GTATGTTGTGCGAATTGATC 3’	L	6492-6474
Final	anti-senso	5’ACGCTTAACAAATAACAACA 3’	Trailer	11904-11924

Para cada amostra, foram adicionados 15µL do RNA extraído ao *mix* para a transcrição reversa contendo 8µL 5X RT Buffer, 6µL do pool de dNTPs na concentração de 10mM, , 5µL do *primer* anti-senso “Final” (Quadro 3) na concentração de 10µM, 400U da enzima transcriptase reversa RevertAid™ Premium Reverse Transcriptase (Fermentas™), 1µL de RNAsin (Invitrogen™) e 35µL de água ultra pura livre de DNase e RNase esterilizada para um volume final de 50µL, realizando-se a transcrição reversa a 42°C/180 minutos seguido de um ciclo de inativação da transcriptase reversa de 70°C/15 minutos.

Após a RT, os cDNAs foram acondicionados em freezers -20 °C até o momento de sua utilização.

3.4 REAÇÃO EM CADEIA PELA POLIMERASE (PCR) DIRIGIDA A EXTREMIDADE 3’ UTR DO RABV - PCR-CONTROLE.

Após a realização da RT para a obtenção do DNA complementar (c-DNA) ao RNA anti genômico do RABV das amostras do quadro 2, foi realizada uma reação de PCR utilizada como controle para a verificação da transcrição reversa dos genomas extraídos, para sua posterior utilização na amplificação do genoma completo do RABV.

Esta reação foi dirigida para amplificação de um fragmento de 1533 pares de base (pb) correspondendo ao início da extremidade 3' UTR (Leader) do genoma viral, o gene codificador da nucleoproteína e o Início do gene codificador da fosfoproteína do RABV. Para esta amplificação foram utilizados os *primers* Início (senso) e 304 (anti-senso) descritos por Campos et al., (2011) e Orciari et al., (2001) respectivamente (Quadro 3).

Esta reação consistiu de 5 µL do c-DNA complementar ao RNA genômico de amostra ao *mix* da PCR contendo 5µL de 10X PCR Buffer (InvitrogenTM), 8 µL do pool de dNTPs a 1,25 mM, 2,5 µL de cada primer a 10 µM, 2,5 µL de MgCl₂ 50 mM, 25µL água ultra pura livre de DNase e RNase esterilizada e 2,5U de Taq DNA polimerase (InvitrogenTM) para um volume final de 50 µL e levados ao termociclador e submetidos ao ciclo descrito no quadro 4.

Quadro 4 – Ciclos de temperaturas utilizados na PCR controle da obtenção do c-DNA genômico dos RABV testados – São Paulo – 2014

Ciclo	Temperatura	Tempo
1 X	94°C Denaturação	5 minutos
35 X	94°C Denaturação	45 segundos
	55°C Hibridação	45 segundos
	72°C Extensão	2 minutos
1 X	72°C Extensão	10 minutos

Os produtos de PCR foram submetidos à eletroforese em gel de agarose a 1% em tampão TBE 1X (0,1 M de Tris, 0,09 M de ácido bórico e 0,001 M de EDTA), contendo brometo de etídeo na proporção de 7µL para cada 100 ml de tampão.

Foram considerados viáveis para a amplificação do genoma completo do RABV os controles que resultaram no fragmento de 1533 pb, o que corresponde ao tamanho do fragmento gerado pela combinação dos *primers* Início/304, demonstrando assim a transcrição reversa do RNA anti genômico total do RABV a partir da extremidade 3' UTR até a extremidade 5' UTR, necessária para a amplificação do genoma completo do RABV.

3.5 PCR DE AMPLICON DE TAMANHO GENOMICO

A padronização desta técnica foi realizada neste trabalho utilizando-se a amostra fixa CVS-11 mantida em camundongos no laboratório de virologia do Instituto Pasteur de São Paulo.

Para esta reação, foram utilizados os cDNAs a partir dos quais se obteve o fragmento de 1533 pb na PCR controle da transcrição reversa do RNA genômico do RABV, descrita no item 3.5.

O protocolo da reação de PCR padronizado para a amplificação do genoma completo do RABV consistiu de: 25 µL de GoTaq® Long PCR Master Mix 2X (Promega), 1 µL dos *primers* Início (senso) e Final (anti- senso) na concentração de 10 µM, 1 µL de c-DNA e 22 µL água ultra pura livre de DNase e RNase esterilizada para um volume final de 50 µL e levados ao termociclador e submetidos ao ciclo descrito no quadro 5.

Quadro 5 – Ciclos de temperaturas utilizados na PCR para a amplificação do genoma do RABV da amostra CVS-11 – São Paulo – 2014

Ciclo	Temperatura	Tempo
1 X	94°C Desnaturação	5 minutos
35 X	94°C Desnaturação	45 segundos
	55°C Hibridação	45 segundos
	68°C Extensão	12 minutos
1 X	68°C Extensão	10 minutos

Os produtos de PCR foram submetidos à eletroforese em gel de agarose a 0,8% em tampão TBE 1X (0,1 M de Tris, 0,09 M de ácido bórico e 0,001 M de EDTA), contendo brometo de etídeo na proporção de 7µL para cada 100 ml de tampão.

Foram consideradas positivas as reações que resultaram no fragmento de aproximadamente 12 kb, o que corresponde ao tamanho do fragmento gerado pela combinação dos primers Início/Final.

3.6 AMPLIFICAÇÃO DO GENOMA DO RABV EM DUAS REAÇÕES DE PCR

As amostras negativas na PCR de tamanho genômico foram amplificadas em duas etapas de PCR, denominadas PCR A (6492 pares de base) e PCR B (7091 pares de base), utilizando-se os *primers* Início/LA3 e VT1551/Final respectivamente (Quadro 3).

Os produtos de PCR foram submetidos à eletroforese em gel de agarose a 0,8% em tampão TBE 1X (0,1 M de Tris, 0,09 M de ácido bórico e 0,001 M de EDTA), contendo brometo de etídeo na proporção de 7µL para cada 100 ml de tampão.

Foram consideradas positivas as reações que resultaram nos fragmentos de aproximadamente 6.5 kb (PCR A) e 7 kb (PCR B).

3.7 PURIFICAÇÃO DAS REAÇÕES DE PCR

A purificação do produto da amplificação dos genomas, bem como das reações de PCR A e B foi realizada a partir do gel de agarose a 1%, após eletroforese com 45µL do produto da PCR, utilizando-se o *kit* QIAquick® Gel Extraction Kit, segundo instruções do fabricante.

3.8 SEQUENCIAMENTO DE DNA PARA A CONFIRMAÇÃO DA AMPLIFICAÇÃO DO GENOMA COMPLETO DO RABV E DAS REAÇÕES DE PCR A E B

Para a realização da técnica de sequenciamento de DNA para a confirmação das amplificações foram utilizados os produtos das reações de PCR obtidas nas 18 amplificações do genoma viral e nas quatro amplificações referentes às quatro amostras amplificadas nas reações de PCR A e B no seguinte protocolo: 4 µL de BigDye 3.1 (Applied Biosystems™), 3,2 pmoles de cada *primer* utilizado na amplificação de cada amplicon (Quadro 3), em reações separadas, 20 ng do DNA alvo e água DNase free q.s.p. para uma reação final de 10 µL, levando-se ao termociclador Mastercycler Gradient (Eppendorf) para 35 ciclos de 96

°C/10 segundos, 50 °C/5 segundos e 60°C/4 minutos, com rampa de 1°C/segundo entre cada temperatura.

A purificação da reação de sequenciamento foi realizada por Sephadex™ G-50 fine (GE healthcare Bio-sciences), em placas com filtro Multiscreen HV (marca) com 96 orifícios. Após a purificação, as sequências foram obtidas em sequenciador automático de DNA ABI-3130 (Applied Biosystems™). Para cada um dos nucleotídeos mostrados nos eletroferogramas gerados para cada uma das reações de sequenciamento foram atribuídos escores através do aplicativo Phred em <http://asparagin.cenargen.embrapa.br/phph/>, sendo utilizadas as posições que apresentaram nucleotídeos com índice Phred maior que 20 (EWING; GREEN, 1998).

Os nucleotídeos com índice Phred igual ou menor a 20 foram conferidos manualmente com o programa Chromas v. 2.23 (© 1998-2002 Technelysium Pty LTD) para a busca por erros de interpretação em cada uma das fitas sequenciadas, sendo as mesmas submetidas ao BLASTn para confirmação do sequenciamento das extremidades 3' e 5' da amostra sequenciada.

Como esta etapa não visa à análise das sequências, mas somente a confirmação do sequenciamento das extremidades 5' e 3' do RABV, não foi necessário um tamanho de sequência pré definido para esta confirmação, somente sua identificação como RABV e a localização no genoma obtido através do aplicativo BLASTn.

3.9 PREPARO DA BIBLIOTECA DE DNA DAS AMOSTRAS E SEQUENCIAMENTO GENÔMICO POR NEXT GENERATION SEQUENCING (NGS) NO SEQUENCIADOR ILLUMINA MISEQ

3.9.1 Diluição das amostras

Após a purificação, os amplificadores genômicos, A e B foram quantificados para a concentração de DNA, através do equipamento Qubit® 2.0 Fluorometer (Invitrogen) e resuspendidos em 25 µl q.s.p. em água DNase RNase free na concentração de 1 ng/µl de DNA no caso dos amplicons genômicos e 0,5 ng/µl no caso dos amplificadores A e B, para a

sua utilização na reação de sequenciamento de DNA pelo método de sequenciamento por síntese na plataforma Illumina MiSeq.

3.9.2 Fragmentação, inserção dos adaptadores e neutralização

Nesta etapa o DNA é tagmentado (etiquetado e fragmentado) pelo transposon Nextera XT transposome, o qual simultaneamente fragmenta a amostra de DNA e adiciona sequências de adaptadores em suas extremidades, que permitirão que estes fragmentos possam ser amplificados por PCR nos passo subsequente.

A fragmentação foi realizada em placas de PCR de 96 poços. Foi utilizado 1 ng das amostras já diluídas, nas quais foram adicionados 10µL de Tagment DNA Buffer e 5µL de Amplicon Tagment Mix. A placa foi centrifugada a 280xg por 1 minuto e depois colocada em um termociclador com o programa: 55°C por 5 minutos e incubação a 10°C. Assim que a temperatura atingiu os 10°C a placa foram adicionados 5µL de Neutralize Tagment Buffer. A placa f centrifugada a 280xg por 1 minuto e depois incubada a temperatura ambiente por 5 minutos.

3.9.3 Amplificação

Nesta etapa o DNA já tagmentado é amplificado por PCR convencional com primers complementares aos adaptadores, ao mesmo tempo que adiciona um par de índice (barcodes) nas extremidades de cada amostra, permitindo sua identificação.

Foram adicionados à placa 15µL de Nextera PCR Master Mix, 5µL de Index 1 Primers e 5µL de Index 2 Primers. A placa é então centrifugada a 280xg por 1 minuto e colocada no termociclador com o seguinte programa: 72°C por 3 minutos, 95°C por 30 segundos seguido de 12 ciclos de 95°C por 10 segundos, 55°C por 30 segundos e 72°C por 30 segundos, uma extensão final a 72°C por 5 minutos e incubação a 10°C.

3.9.4 Purificação

Nesta etapa são utilizadas AMPure XP beads para purificar a biblioteca de DNA e realizar uma seleção de tamanho que remove fragmentos muito curtos da população de bibliotecas.

A placa na qual estão as bibliotecas foi centrifugada por 1 minuto a 280 xg a 20 °C. Em uma nova placa de PCR de 96 poços foram transferidos 50 µL de cada biblioteca (uma biblioteca em cada poço) e adicionados 30 µL de AMPure XP beads, sendo a placa então incubada a temperatura ambiente por 5 minutos.

Após a incubação a placa foi colocada na estante magnética por 2 minutos ou até o clareamento do sobrenadante o qual é aspirado. Ainda na estante magnética, a placa então foi lavada com 200 µL de Etanol 80% e incubada por 30 segundos, sendo o sobrenadante descartado após esta etapa e o procedimento de lavagem repetido. Após esta etapa, o sobrenadante foi aspirado e a placa foi seca por 15 minutos a temperatura ambiente ainda acoplada à placa magnética.

Após sua secagem, a placa foi retirada da estante magnética e foram adicionados 52,5 µL de Resuspension Buffer (RSB) em cada poço, misturados 10 vezes através de pipetagem, sendo a placa deixada em temperatura ambiente por 2 minutos. A placa foi então colocada novamente na estante magnética por 2 minutos ou até o clareamento do sobrenadante, sendo que ao final desta etapa, 50 µL de cada amostra foram transferidos para uma nova placa de PCR de 96 poços.

3.9.5 Normalização

Após a etapa de amplificação, a biblioteca foi purificada para retirar excessos de sais, enzimas, adaptadores e outros reagentes. 20 µL do produto amplificado foram transferidos para uma nova placa. Em um tubo tipo falcon foram adicionados 4,4mL de LNA1 (Library Normalization Additives 1) e 800µL de LNB1 (Library Normalization Beads 1). A solução foi homogeneizada por inversão. 45µL da mistura foram adicionadas a cada poço da placa

contendo os 20µL do produto amplificado. A placa foi então agitada a 1.800 rpm por 30 minutos. Após a agitação a placa foi colocada em uma estante magnética e 80µL do sobrenadante são removidos com o cuidado de não tocar no pellet formado. A placa foi então removida da estante e o pellet lavado com 45µL de LNW1 (Library Normalization Wash 1). A placa foi novamente agitada a 1.800 rpm por 5 minutos e depois colocada na estante magnética para a remoção do sobrenadante. Os processos de lavagem, agitação e retirada do sobrenadante foram repetidos sob as mesmas condições. Em seguida a placa foi removida da estante e adicionados 30µl de NaOH a 0,1 N sendo a mesma é novamente agitada a 1.800 rpm por 5 minutos. Após a agitação a placa foi novamente colocada na estante e 30µl do sobrenadante foram transferidos para uma placa limpa contendo 30µl de LNS1 (Library Normalization Storage Buffer1).

3.10 AMPLIFICAÇÃO DOS CLUSTERS E SEQUENCIAMENTO POR SINTESE NO SEQUENCIADOR ILLUMINA MISEQ

5µl de cada poço foram transferidos para um único tubo e em seguida 24µl desse tubo foram transferidos para um segundo tubo contendo 576µl de HT1 (Hybridization buffer 1), agitado e encubado a 96 C por 2 minutos e logo em seguida colocado em um banho de gelo. Em seguida adicionou-se todo o conteúdo do tubo no cartucho de sequenciamento Miseq já descongelado no reservatório de carregamento de amostras.

3.11 EDIÇÃO DE SEQUÊNCIAS

Os reads sequenciados na plataforma Illumina MiSeq foram montados sem o uso de genomas de referência (montagem *de novo*) no programa CLC Genomics Workbench 6. Para os genomas em que os maiores *contigs* montados tinham um tamanho inferior ao esperado de acordo com o tamanho do fragmento de PCR sequenciado, estes contigs foram estendidos por mapeamento dos reads à referência com 95% de similaridade e sem abertura de gaps, ou pelo

método de montagem *de novo* com o mesmo critério de similaridade, no programa Geneious R6. Tanto os contigs montados *de novo* quanto os estendidos tiveram os reads novamente mapeados à referência e foram avaliados quanto à variabilidade com base no número e qualidade de reads através da função "Quality-based Variant Detection" do programa CLC.

As sequências de baixa qualidade foram retiradas da análise e os parâmetros utilizados foram: qualidade de base *phred* igual a 20, cobertura mínima de 100 leituras por sítio, e penalidade para regiões homopoliméricas (redução de qualidade de 30% para cada base adicional).

Após a edição, as regiões das sequências de RABV correspondentes aos *primers Início* e *Final* foram removidas.

3.12 ANÁLISE DESCRITIVA DAS SEQUÊNCIAS OBTIDAS NESTA PESQUISA

A análise descritiva intra-cluster das sequências correspondentes às linhagens brasileiras de RABV obtidas nesta pesquisa, foi realizada através do software BioEdit v. 5.0.9 (HALL, 1999).

A amostra CVS-11, utilizada como controle nesta pesquisa, foi comparada com a amostra CVS-11 (acession number GQ918139.1), mantida pelo National Institute for the Control of Pharmaceutical and Biological Products da China, para a procura de substituições e também como forma de controle da qualidade da sequência obtida.

3.13 CONJUNTO DE DADOS PARA AS ANÁLISES

Para a realização das análises descritas a seguir, foi utilizado um data set de 138 sequências de genoma completos de RABV extraídas do GenBank e 21 das 22 sequências obtidas neste estudo, contemplando um alinhamento total de 159 sequências com um tamanho variando entre 11719 e 11742 nucleotídeos, do nucleotídeo 59 ao 11.801 em relação a amostra fixa PV, utilizada arbitrariamente como régua para a edição do alinhamento. Esta região contempla desde o Início do RNA mensageiro do gene N, até o códon de parada do gene L.

Visando enriquecer as análises, selecionaram-se entre as sequências de RABV do GenBank aquelas com mais de 10000 nucleotídeos, cujo ano de isolamento, país e hospedeiro estavam acessíveis. Para tal, foi adaptado um programa em linguagem Perl para se automatizarem as buscas.

Para a realização dos estudos evolutivos, não foi utilizada nenhuma amostra fixa mantida em laboratório. Por este motivo a amostra CVS sequenciada neste estudo não foi utilizada nas análises filogenéticas.

O tamanho final dos alinhamentos, incluindo os *gaps* gerados pela diversidade entre as sequências foi de 11.785 posições.

As análises foram realizadas utilizando-se o alinhamento completo descrito acima, bem como os alinhamentos dos genes separadamente e Pseudo gene.

3.14 ANÁLISE DE RECOMBINAÇÃO

Inicialmente, as sequências foram alinhadas com o programa MUSCLE v3.7 (EDGAR, 2004), em seguida o alinhamento foi editado manualmente com Bioedit v. 5.0.9 (HALL, 1999). A detecção de possíveis eventos de recombinação homóloga, identificação de sequências recombinantes e localização de pontos de recombinação foram investigadas por meio dos métodos RDP, GenConv, Chimaera, MaxChi, Bootscan, SiScan e 3Seq implementados no programa RDP4 beta48 (MARTIN et al., 2010), utilizando-se intervalo de confiança de 95% e correção múltipla de Bonferroni para se evitar resultados falsos positivos.

3.15 ANÁLISE DE SINAL FILOGENÉTICO

Para se investigar o conteúdo de sinal filogenético dos genes e pseudo gene do RABV separadamente, foi utilizado o algoritmo de mapeamento de verossimilhança (STRIMMER; HAESELER, VON, 1997) implementado em TREE-PUZZLE v 5.2 (SCHMIDT et al., 2002) que quantifica os quartetos filogenéticos bem resolvidos nos bancos de dados. Esta análise

utilizou todas as 159 (138 sequências oriundas do Genbank e 21 sequenciadas neste trabalho) sequências correspondentes aos ciclos aéreo (100 sequências) e terrestre do RABV (59 sequências).

3.16 ANÁLISE DE HETEROTAQUIA

Heterotaquia é o fenômeno no qual as taxa de evolução de determinados sítios ou regiões codificantes podem variar em uma dada filogenia. A Heterotaquia descreve uma inferência, a partir de um alinhamento múltiplo, de que existem diferenças nas taxas evolutivas para determinadas regiões deste alinhamento, sendo estas diferenças linhagens-específicas, ou seja, estas diferentes linhagens estão sob diferentes relógios moleculares. Estas diferentes taxas de evolução podem ser relacionadas a diferentes processos seletivos atuando nas sequências sob heterotaquia (LOCKHART et al., 2006).

Esta análise utilizou todas as 159 sequências correspondentes aos ciclos aéreo e terrestre do RABV. A análise da heterotaquia foi realizada utilizando-se o programa PAML v4.6 (YANG, 2007), para se testar as seguintes premissas: (i) as linhagens virais de ambos os ciclos dos RABV (ciclo aéreo mantido principalmente por quiróptero e ciclo terrestre mantido principalmente pelo cão) estão sob o mesmo relógio molecular e (ii) as linhagens pertencentes a cada ciclo estão sob relógios distintos. Para tal, foram utilizadas 100 árvores filogenéticas de máxima verossimilhança inferidas com o programa GARLI v 2.0 (ZWICKL, 2006) utilizando-se aquela com o maior valor de verossimilhança.

Para se testar a significância da diferença entre os valores de verossimilhança (V) de cada modelo estudado, aplicou-se um teste de diferença relativas de verossimilhança (LRT, da sigla em inglês), segundo a equação abaixo.

$LRT = 2 \times (L1 - L2)$, onde L1 é igual à verossimilhança do modelo 1 e L2 a verossimilhança do modelo 2. Nesse caso, os graus de liberdade empregados foram iguais a 157 que equivalem ao número de taxa subtraído de duas unidades (N-2). O nível de significância estabelecido foi de 0,05, cujo valor crítico é de 206,390, em uma tabela de qui-quadrado.

O resultado desta análise foi de fundamental importância para a realização de diversas as análises evolutivas do RABV contempladas nesta pesquisa, uma vez que foi comprovada a existência de relógios moleculares distintos para cada ciclo, seria necessária a análise dos

mesmos separadamente para conhecermos quais os genes ou regiões do genoma do RABV responsáveis por este fenômeno, para que fossem retirados das análises para analisar conjuntamente os ciclos aéreo e terrestre do RABV.

3.17 ESTIMATIVAS DA TAXA DE SUBSTITUIÇÃO DE NUCLEOTÍDEOS POR SÍTIO POR ANO, TEMPO PARA O ANCESTRAL COMUM MAIS RECENTE (TACMR) E ESTIMATIVAS DE FILOGENIAS

Devido ao resultado positivo para a existência de heterotaquia (item 4.5), foi observado que não seria possível a utilização conjunta dos ciclos aéreos e terrestre do RABV para a realização das estimativas da taxa de substituição de nucleotídeos e tempo para o ancestral comum mais recente, enquanto não se soubesse quais entre os cinco genes do RABV seriam os ideais para esta análise conjunta.

Sendo assim, estas estimativas foram feitas inicialmente separadamente para cada um dos cinco genes e região intergênica G-L nos dois ciclos, e após a análise dos resultados, utilizamos os genes com taxas de substituição similares nos dois ciclos para então estimarmos o tempo de ancestral comum mais recente (TACMR) entre eles.

3.17.1 Estimativa da taxa de substituição de nucleotídeos por sítio por ano e tempo para o ancestral comum mais recente (TACMR) calculadas separadamente para os ciclos aéreo e terrestre do RABV.

As taxas de substituição por sítio por ano foram calculadas para cada gene separadamente, através do programa BEAST v 1.7.4.(DRUMMOND et al., 2012). Anteriormente às análises com o BEAST, foram estimadas taxas iniciais de substituição para serem utilizadas como priori nas análises Bayesianas, utilizando o programa Parth-O-Gen (<http://tree.bio.ed.ac.uk/software/pathogen>).

Árvores filogenéticas com a máxima credibilidade de clados (MCC) foram inferidas para cada um dos cinco genes do RABV usando uma abordagem baseada em Markov Chain

Monte Carlo (MCMC) implementada em BEAST v1.7.4 (DRUMMOND et al 2012) empregando o modelo para sequências codificantes SRD6 (Shapiro, Rambaut, & Drummond, 2003; Wertheim & Kosakovsky Pond, 2011) e relógio molecular relaxado não correlato, utilizando uma distribuição probabilística lognormal (DRUMMOND et al., 2006) com o μ (taxa de substituição de nucleotídeos) previamente estimado. A convergência do MCMC foi obtida após duas execuções independentes com 50 milhões de gerações cada uma, amostrando a cada 5 mil árvores, que foram suficientes para se obter uma amostragem apropriada da estacionariedade do MCMC que foi inspecionada com auxílio do programa Tracer v 1.5 (disponível em <http://tree.bio.ed.ac.uk/software/>), considerando-se suficiente quando os tamanhos amostrais efetivos (ESS) dos parâmetros atingiram valores superiores a 200. Para o pseudo gene foi utilizado o modelo evolutivo GTR+ Γ +I.

3.17.2 Estimativa da taxa de substituição de nucleotídeos por sítio por ano e tempo para o ancestral comum mais recente (TACMR) calculadas em conjunto para os ciclos aéreo e terrestre do RABV

Após a realização das análises descritas nos itens 3.15 e 3.17.1 chegamos à conclusão de que a melhor maneira de estimar o provável tempo da separação entre os ciclos aéreo e terrestre do RABV seria através da utilização dos genes G e L concatenados, pois os genes N, P e M, que estão sob heterotaquia nos ciclos aéreo e terrestre, apresentaram diferentes taxas de substituição de nucleotídeos para os ciclos aéreo e terrestre do RABV, enquanto os genes G e L apresentaram taxas de substituições de nucleotídeos similares para os dois ciclos.

Árvores filogenéticas com a máxima credibilidade de clados (MCC) utilizando o gene G e L concatenados foram inferidas usando uma abordagem baseada em MCMC implementada em BEAST v1.7.4 (DRUMMOND et al 2012) empregando SRD6 e relógio molecular relaxado não correlato, utilizando uma distribuição probabilística lognormal (DRUMMOND et al. 2006) com o μ previamente estimado, com base nos resultados obtidos na análise separadas de cada ciclo. A convergência do MCMC foi obtida após quatro execuções independentes com 100 milhões de gerações cada uma, amostrando a cada 10 mil árvores, que foram suficientes para se obter uma amostragem apropriada da estacionariedade do MCMC que foi inspecionada com auxílio do programa Tracer v 1.5 (disponível em

<http://tree.bio.ed.ac.uk/software/>), considerando-se suficiente quando os tamanhos amostrais efetivos (ESS) dos parâmetros atingiram valores superiores a 200.

Para testar a recuperação dos mesmos agrupamento genéticos a partir das sequências genômicas, foi gerada uma árvores de máxima verossimilhança utilizando o programa GARLI v2.0 (BAZINET; ZWICKL; CUMMINGS, 2014) que emprega um algoritmo genético estocástico para estimar simultaneamente a melhor topologia, comprimentos de ramos e parâmetros do modelo de substituição empregado que maximizam o valor logarítmico da Verossimilhança (lnL), dez replicações independentes foram executadas sob o modelo de substituição *general time reversible* (GTR) com taxa de variação gama e uma proporção de sítios invariantes (GTR+ Γ +I). A replicação com lnL mais alto foi selecionada. Objetivando-se avaliar o suporte estatístico dos ramos das árvores, foram empregadas mil amostragens não paramétricas de *bootstrap*.

3.18 ANÁLISE DO REGIME DE SELEÇÃO

Pretendendo-se inferir as pressões de seleção agindo sobre os genes analisados, estimou-se a diferença entre as taxas de substituição não sinônimas (dN) e sinônimas (dS) por sítio codificante usando-se algoritmo de contagem ancestral de única verossimilhança (SLAC) disponível em HyPhy v2.11beta (POND; FROST; MUSE, 2005), assumindo um nível de confiança de 95% ($\alpha = 0,05$). Valores da diferença ω (onde ω representa dN-dS) maiores que zero são indicativos de seleção direcional positiva ($\omega > 0$), os valores iguais a zero indicam seleção neutra, enquanto valores menores que zero indicam seleção negativa purificadora. Esta análise foi realizada para cada um dos cinco genes do RABV, separadamente para os ciclos aéreo e terrestre.

A significância estatística das diferenças encontradas para ω , com relação aos 5 genes dos RABV dos ciclos aéreo e terrestre foi calculada através do teste do sinal ranqueado de Wilcoxon.

Para o gene P, foram utilizados 879 nucleotídeos, relativos a 293 aminoácidos devido ao fato este ser o tamanho de algumas sequências do gene P utilizadas para as análises. Normamente o tamanho do gene P nos RABV é de 891 nucleotídeos, responsáveis pela tradução de 297 aminoácidos.

Para o gene N foram utilizados 1497 nucleotídeos, relativos a 449 aminoácidos, uma vez que a amostra IP 1770/12, sequenciada nesta pesquisa, apresentou um códon de parada TGA na posição 450 da nucleoproteína, impossibilitando o uso deste códon na análise.

3.19 ANÁLISE DO USO PREFERENCIAL DE CÓDONS

Para esta análise foram analisados o RSCU (uso relativo de códons sinônimos), o Nc (número efetivo de códons) e o CAI (índice de adaptação do códon), os quais foram calculados para todos os genes do RABV, separadamente para os ciclos aéreo e terrestre, com o objetivo de encontrar diferentes padrões de uso preferencial de códons sinônimos para os dois ciclos.

O uso relativo de códons sinônimos (RSCU) representa a relação entre o número de códons sinônimos utilizados e o número esperado de códons sinônimos se sua utilização fosse randômica., sendo que valores de $RSCU > 1$ indicam o uso de um códon sinônimo como preferencial; $RSCU = 1$ indica a ausência de preferência e $RSCU < 1$ indicam códons pouco utilizados ou preteridos. Esta análise foi realizada separadamente para o ciclo aéreo e ciclo terrestre, a fim de evidenciar algum padrão específico de RSCU em determinado aminoácidos para cada um dos ciclos, utilizando-se o Sequir no pacote R (CHARIF et al., 2005).

Depois de estimados os RSCUs para cada sequência em cada ciclo, os códons foram categorizados em preferidos ($RSCU > 1$) e preteridos ($RSCU \leq 1$) e então aplicado o teste exato de Fisher com nível de significância ($\alpha = 0,05$) nos códons preferidos para cada aminoácido entre os ciclos, para se testar a diferença estatisticamente significativa para o uso de cada códon em um mesmo gene para cada ciclo.

O número efetivo de códons (ENC) é a mensuração do numero total de diferentes códons presentes em uma sequência e mostra o viés do uso de códons sinônimos para determinado aminoácido. O ENC foi calculado para todas as sequências, analisando-se separadamente cada gene nos ciclos aéreo e terrestre do RABV, utilizando-se o software ACUA 1.0 (VETRIVEL; ARUNKUMAR; DORAIARAJ, 2007).

O CAI representa a média geométrica da adaptação relativa de cada códon calculada utilizando como parâmetro um set de códons otimamente traduzidos em genes de alta

expressão e varia de 0 a 1. Os valores de CAI estão diretamente relacionados a otimização de tradução dos códons contidos no gene ou região analisada. Valores próximos de 1 indicam alta adaptação na tradução dos códons do gene ou região analisada em relação aos genes de alta expressão do organismo utilizado como referência. Esta análise foi realizada separadamente para cada gene do RABV nos ciclos aéreo e terrestre, através do software ACUA 1.0 (VETRIVEL et al., 2007) utilizando-se como referência genes de alta expressão de *Mus musculus*.

3.20 DIVERSIDADE GENÉTICA CALCULADA ATRAVÉS DA DISTÂNCIA PATRÍSTICA E SUA COMPARAÇÃO COM OS RESULTADOS DE DISTÂNCIA GENÉTICA E IDENTIDADE GENÉTICA.

A distância patrística é a soma dos comprimentos de ramos que conectam dois nós terminais ou taxas em uma árvore filogenética. Uma matriz de distância patrística calculada a partir de uma árvore filogenética para todos os taxas par a par representa a divergência genética ou a variação filogenética representada pelo conjunto de sequências contidos nesta árvore filogenética. Os resultados obtidos através do cálculo da distância patrística podem ser utilizados para análise de variabilidade genética ou utilizados para comparação com outras medidas de divergência genética (FOURMENT; GIBBS, 2006).

O cálculo da distância patrística nos ciclos aéreo e terrestre do RABV foi realizado em conjunto e separadamente para cada um dos 5 genes virais utilizando-se o software PATRISTICv1.0. (FOURMENT; GIBBS, 2006).

A significância estatística entre as diferenças de distância patrística encontradas na análise em separado entre os ciclos aéreo e terrestre do RABV em relação aos mesmos pares de genes (ex: N aéreo X N terrestre) entre 5 genes analisados foi calculada utilizando-se o teste do sinal ranqueado de Wilcoxon ($\alpha = 0,05$).

A significância estatística entre as diferenças de distância patrística entre cada um dos cinco genes do RABV na análise em conjunto dos ciclos aéreos e terrestre foi calculada utilizando-se o Kruskal-Wallis rank sum test ($\alpha = 0,05$).

Para o cálculo da distância patrística foram utilizadas as árvores filogenéticas de máxima verossimilhança inferidas separadamente para cada um dos 5 genes do RABV nos

ciclos aéreo e terrestre, utilizando-se aquelas com maiores valores de verossimilhança obtidas após 100 corridas independentes com o programa GARLI v 2.0 (BAZINET et al., 2014).

As identidades genéticas mínima, máxima e média para os genes N, P, M, G e L foram calculadas em conjunto e separadamente como ciclo aéreo e terrestre com o programa Excel (©1985-2003 Microsoft Corporation) a partir das matrizes de identidades calculadas com o programa Bioedit (HALL, 1999).

As distâncias genéticas médias e os erros padrões para os genes N, P, M, G e L foram calculadas em conjunto e separadamente como ciclo aéreo e terrestre com o software MEGA 5 (© 1993 – 2011 TAMURA et al.) usando-se o modelo evolutivo de MCL.

Para testar a correlação entre os resultados obtidos a partir dos cálculos da distância patrística, identidade genética e distância genética para os ciclos aéreo e terrestre separadamente e em conjunto utilizou-se o teste de regressão linear (r^2) par a par para cada método através do programa Excel.

Para o gene P, foram utilizados 879 nucleotídeos, relativos a 293 aminoácidos devido ao fato ser este o tamanho de algumas sequências do gene P utilizadas para as análises. Normamente o tamanho do gene P nos RABV é de 891 nucleotídeos, responsáveis pela tradução de 297 aminoácidos.

Para o gene N foram utilizados 1497 nucleotídeos, relativos a 449 aminoácidos, uma vez que a amostra IP 1770/12, sequenciada nesta pesquisa, apresentou um códon de parada TGA na posição 450 da nucleoproteína, impossibilitando o uso deste códon na análise.

3.21 TESTES ESTATÍSTICOS

Todos os testes estatísticos foram feitos no programa R (<http://www.r-project.org/>), as variáveis discretas foram comparadas utilizando-se o teste exato de Fisher, as variáveis contínuas foram analisadas com o teste do sinal ranqueado de Wilcoxon e as comparações múltiplas foram realizadas com o teste Kruskal-Wallis.

4 RESULTADOS

4.1 RT- PCR PARA AMPLIFICAÇÃO DO GENOMA COMPLETO DO RABV

Estas reações foram padronizadas nesta pesquisa utilizando-se RNA extraído de SNC de camundongos inoculados com a amostra CVS-11, de acordo como descrito em material e métodos nos itens 3.4, 3.5 e 3.6.

Após sua padronização, esta reação foi testada para a amplificação de 22 amostras de SNC de camundongos inoculados com diferentes linhagens brasileiras de RABV (Quadro 3)

Como resultado foram obtidas 18 reações de PCR com bandas de aproximadamente 12 kpb (Figura 1), tamanho aproximado do genoma do RABV.

Figura 1 – Visualização em gel de agarose a 1% da banda de aproximadamente 12 kpb resultante da amplificação do genoma completo da amostra CVS-11 - São Paulo - 2014

Fonte: Instituto Pasteur

4.2 AMPLIFICAÇÃO DO GENOMA DO RABV EM DUAS REAÇÕES DE PCR

Para as amostras 350/10, 3208/06, 5402/07 e 6634/08, a reação de PCR de tamanho genômico não funcionou. Sendo assim, estas amostras foram amplificadas em duas reações de PCR (PCR A e B), como descrito no item 3.8.

4.3 SEQUENCIAMENTO DE DNA PARA A CONFIRMAÇÃO DA AMPLIFICAÇÃO DO GENOMA COMPLETO DOS RABV AMPLIFICADOS E DAS REAÇÕES DE PCR A E B

Para a demonstração da amplificação dos genomas completos, bem como das reações de PCR A e B, os amplicons obtidos foram purificados e suas extremidades 3' e 5' foram submetidas à técnica de sequenciamento de DNA com os *primers* utilizados para a sua amplificação (Quadro 3), de acordo com o protocolo descrito no item 3.7.

Após edição, foram obtidos para cada amostra duas sequências de DNA referentes a cada *primer* utilizado, que após serem editadas e analisadas mostraram-se correspondentes às extremidades 3' UTR e 5' UTR do RABV, confirmando assim a amplificação do genoma completo do RABV das amostras utilizadas nesta pesquisa.

4.4 AMPLIFICAÇÃO DOS CLUSTERS E SEQUENCIAMENTOS POR SÍNTESE NA PLATAFORMA ILLUMINA MISEQ.

As 22 amostras de DNA provenientes das amplificações de genomas de RABV foram sequenciadas com sucesso, sendo que após sua montagem e edição foram obtidas sequências de DNA correspondentes ao nucleotídeo 23 até o nucleotídeo 11911 em relação à amostra PV (M13215.1). Esta região corresponde ao primeiro nucleotídeo após a região de hibridação para o *primer* Início e o nucleotídeo anterior a região de hibridação para o *primer* Final, já que

hibridações de baixa especificidade de nucleotídeos não complementares nestas regiões podem acontecer o que não garante a fidelidade das sequências geradas para estas posições.

Apenas para a amostra CVS-11 não se obteve sequenciamento viável da região 5' UTR (Quadro 6).

Quadro 6 – Tamanho das sequências de DNA obtidas e média de leituras de cada nucleotídeo para cada amostra - São Paulo - 2014

Amostra	Linhagem	Pares de base	Média de leituras
CVS-11	Fixa	11835 nucleotídeos	544,92
1770/12	<i>Callithrix jacchus</i>	11879 nucleotídeos	1.289,94
5402/07	<i>Callithrix jacchus</i>	11878 nucleotídeos	6.176,29
6634/08	<i>Callithrix jacchus</i>	11878 nucleotídeos	2.859,98
7841/09	<i>Cerdocyon thous</i>	11877 nucleotídeos	3.898,19
4005/10	<i>D. rotundus</i>	11881 nucleotídeos	1.803,01
2990/13	<i>D. rotundus</i>	11879 nucleotídeos	31.051,06
2991/13	<i>D. rotundus</i>	11879 nucleotídeos	1.842,13
2992/13	<i>D. rotundus</i>	11879 nucleotídeos	2.503,75
230/10	<i>Eptesicus 1</i>	11881 nucleotídeos	6.926,89
346/10	<i>Eptesicus 1</i>	11881 nucleotídeos	3.329,40
512/09	<i>Eptesicus 1</i>	11881 nucleotídeos	1.245,95
3208/06	<i>Eptesicus 2</i>	11882 nucleotídeos	9.439,26
163/10	<i>Myotis</i>	11879 nucleotídeos	1.737,81
497/09	<i>Myotis</i>	11880 nucleotídeos	1.775,04
1400/10	<i>Myotis</i>	11880 nucleotídeos	12.341,29
412/10	<i>Nyctinomops</i>	11886 nucleotídeos	1.792,87
542/10	<i>Nyctinomops</i>	11886 nucleotídeos	38.613,78
4431/10	<i>Nyctinomops</i>	11886 nucleotídeos	1.195,17
350/10	<i>Nyctinomops</i>	11886 nucleotídeos	54.458,41
1586/10	<i>Tadarida brasiliensis</i>	11879 nucleotídeos	5.074,54
3176/09	<i>Tadarida brasiliensis</i>	11879 nucleotídeos	4.012,66

4.5 ANÁLISE DESCRITIVA DAS SEQUÊNCIAS OBTIDAS NESTA PESQUISA.

Após sua edição, as sequências agrupadas em cada uma das oito linhagens de RABV aqui estudadas foram alinhadas entre si, para análise intraespecífica.

Quanto ao número de nucleotídeos correspondentes os genes N, P, M, G e L, todas as amostras apresentaram o número de nucleotídeos e aminoácidos esperados para o RABV. A única exceção foi encontrada na amostra 1770/12, isolada da espécie *Callithrix jacchus*, que embora tenha apresentado 1350 nucleotídeos em seu gene N, apresentou 449 aminoácidos na proteína traduzida por ele, enquanto o esperado seriam 450 aminoácidos. Isto ocorreu devido a uma substituição no último códon deste gene para o códon de parada TGA (Figura 2).

Figura 2 – Mutação de nucleotídeo e substituição do último aminoácido da nucleoproteína da amostra IP 1770/12 - São Paulo - 2014

Entre as amostras representantes das linhagens *C. jacchus*, *D. rotundus* e *Myotis* Brasil sequenciadas nesta pesquisa também foram encontradas inserções de nucleotídeos em regiões não codificantes as quais estão descritas na figuras 3 a 6. Esta análise foi realizada com as sequências de DNA sem as regiões correspondentes aos primers Início e Final (nucleotídeo 23 ao 11911 da amostra PV M13215.1).

Figura 3 – inserção de adenina na posição 5332 da amostra IP 1770/12 (linhagem *C. jacchus*) em relação a amostras da mesma linhagem, entre os nt 5363-5364 em relação a amostra PV M13215 - São Paulo - 2014

Figura 4 – Inserção de timina na posição 2430 amostra IP 4005/12 (linhagem *D. rotundus*) em relação a amostras da mesma linhagem na posição 2452 da amostra PV M13215 - São Paulo - 2014


```

          10      20      30      40      50
      . . . . | . . . . | . . . . | . . . . | . . . . |
M13215.1  GAACCTCTCCACTCAGTCCCCTCTAGACAATAAAGTCCGAGATGTCCTAAA
IP 2992/13 ATACCTCCAGGCTCAACCACACTATGCAGTCAGATCTAATCT-GCTTCAA
IP 2990/13 ATATGTCCAGGCTCAACTATACTATGCAGTTAGATCTAACTT-GCTTCAA
IP 2991/13 ATACCTCCAGGCTCAACCACACTATGCAGTCAGATCTAATCT-GCTTCAA
IP 4005/12 ATACATCCAGGCTCAACTATATTATGCAGTCAGATCTAATTCTGCTTCAA

```

Figura 5 – Inserção de adenina na posição 5334 da amostra IP 4005/12 (linhagem *D. rotundus*) em relação a amostras da mesma linhagem, entre os nucleotídeos 5363-5364 da amostra PV M13215 - São Paulo - 2014

```

          10      20      30      40      50
      . . . . | . . . . | . . . . | . . . . | . . . . |
M13215.1  ATTATAAAGTGCTGGGTCACTAAGCTTTTCAGTCGAGAAAAAAA-CAGT
IP 2992/13 ACGACAAATGGATGAACCATCCAAGCCTTTTCAGTCAAGAAAAAAA-CTGT
IP 2990/13 ACTACAAAGGGATGGACCATCCAAGCCTTTCAATCAAGAAAAAAA-CTGT
IP 2991/13 ACGACAAATGGATGAACCATCCAAGCCTTTTCAGTCAAGAAAAAAA-CTGT
IP 4005/12 ACTACAAAGGGATGGACCATCCAAGCCTTTCAATCAAGAAAAAAA-CTGT

```

Figura 6 – Inserção de adenina na posição 2454 da amostra IP 1400/10 (linhagem *Myotis*) em relação a amostras da mesma linhagem, entre os nucleotídeos 2475-2476 da amostra PV M13215 - São Paulo - 2014

```

          10      20      30      40      50
      . . . . | . . . . | . . . . | . . . . | . . . . |
M13215.1  CTAGACAATAAAGTCCGAGATGTCTCTAAAGTCAACATGAAAAAAA-CAGG
IP 1400/10 TTCTGTAATCAAATCTGACCTGCCCAAAATCCAACATGAAAAAAAACAGG
IP 163/10  TTCTGTAATCAAATCTGACCTGCCCAAAATCTAACATGAAAAAAA-CAGG
IP 497/09  CTATGCAATCAAATCTGACCTGCCCAAAATTTAACATGAAAAAAA-CAGG

```

Como descrito por Finke et al. (2000), observou-se nas amostras sequenciadas nesta pesquisas, bem como nas outras sequências de RABV utilizadas no alinhamento completo deste trabalho, o seguinte tamanho das 4 regiões intergênicas do RABV: N/P 2 nucleotídeos, P/M 5 nucleotídeos, M/G 5 nucleotídeos e G/L 24 nucleotídeos.

No alinhamento completo aqui utilizado (159 sequências), o qual utilizou apenas sequencias selvagens, na posição 333 da glicoproteína madura do RABV, o aminoácido encontrado para a maioria das amostras foi a arginina (145 amostras), seguidos de K (5 amostras), Q (4 amostras), N (3 amostras) e H (2 amostras), sendo que a fixação do aminoácido N nesta posição foi exclusiva para os RABV da linhagem *E. furinalis* I enquanto a substituição K foi exclusiva para o complexo de linhagens mantidas por diferentes espécies do gênero *Lasiurus* nos EUA.

A amostra CVS-11, utilizada como controle nesta pesquisa, foi comparada com a amostra CVS-11 (acession number GQ918139.1), mantida pelo National Institute for the Control of Pharmaceutical and Biological Products da China, para a procura de substituições

de nucleotídeos e aminoácidos e também como forma de controle da qualidade da sequência obtida.

Esta análise foi realizada a partir do alinhamento da sequência de 11835 nucleotídeos obtidos a partir do sequenciamento da amostra CVS-11 utilizada nesta pesquisa com a amostra GQ918139.1, correspondendo do nucleotídeo 23 ao 11857 da amostra CVS-11.

A identidade entre nucleotídeos e aminoácidos foi de 99,9%, sendo que para o cálculo da identidade de aminoácidos, as regiões codificantes foram colocadas em passo de leitura.

Foram encontradas quatro substituições de nucleotídeos entre as duas amostras, sendo todas elas transições, três não sinônimas e uma sinônima, descritas no quadro 7.

Quadro 7 – Mutações encontradas entre as amostras CVS-11 GQ918139.1 e a amostra CVS-11 Pasteur - São Paulo - 2014

Nucleotídeo	Mutação	Códon	Aminoácido	TIPO	POSIÇÃO
4541	A ↔ G	AAA ↔ GAA	K ↔ E	TRANSIÇÃO NS	AMINOÁCIDO 408 PROTEÍNA G
4869	G ↔ A	AGG ↔ AAG	R ↔ K	TRANSIÇÃO NS	AMINOÁCIDO 517 PROTEÍNA G
7317	C ↔ T	TTT ↔ TTC	F	TRANSIÇÃO S	AMINOÁCIDO 635 PROTEÍNA L
10174	A ↔ G	AAC ↔ GAC	N ↔ D	TRANSIÇÃO NS	AMINOÁCIDO 1588 PROTEÍNA L

4.6 ANÁLISE DE RECOMBINAÇÃO

Sequências recombinantes potencialmente causam erros sistemáticos em filogenias (SCHIERUP; HEIN, 2000; POSADA; CRANDALL, 2002), portanto foi estabelecida a retirada destas sequências das análises filogenéticas subsequentes.

Não foram encontradas evidências de recombinação entre as amostras de RABV utilizadas para esta análise, sendo assim não foi necessário remover nenhuma das sequências da análise.

4.7 ANÁLISE DE HETEROTAQUIA

Inicialmente, encontrou-se que a hipótese com um relógio molecular global foi menos favorecida que a hipótese com relógios locais. Esse achado foi inferido após se observar um maior valor de verossimilhança para o primeiro modelo, sendo significativa a diferença entre os valores obtidos para cada modelo ($LRT=294,948$), considerando o valor crítico em uma tabela de qui-quadrado 206,39.

Uma vez que um único relógio molecular poderia não ser aplicável a ambos os ciclos devido à existência da heterotaquia, as análises subsequentes foram realizadas separadamente para cada ciclo a fim de encontrarem evidências da heterotaquia, inferir os genes provavelmente responsáveis por este fenômeno para posteriormente realizar as análises filogenéticas em conjunto para os ciclos aéreo e terrestre do RABV utilizando o(s) gene(s) mais adequado para esta finalidade.

4.8 ESTIMATIVAS DA TAXA DE SUBSTITUIÇÃO DE NUCLEOTÍDEOS POR SÍTIO POR ANO E TEMPO DE ANCESTRAL COMUM MAIS RECENTE E ESTIMATIVAS DE FILOGENIA

Primeiramente, esta análise foi realizada separadamente para todos os cinco genes do RABV, estudando-se as amostras dos ciclos aéreo e terrestre independentemente, sendo seus resultados descritos no quadro 8.

Após a análise independente dos ciclos aéreo e terrestre para cada gene, os genes N, M e P apresentaram taxa de substituição de nucleotídeos diferentes em cada ciclo, sendo assim os genes responsáveis pela heterotaquia existente entre os ciclos do RABV.

Enquanto a proteína N apresentou uma taxa de substituição no ciclo terrestre aproximadamente 2 vezes maior em relação ao ciclo aéreo, as proteínas P e M apresentaram uma taxa de substituição aproximadamente 2 vezes maior no ciclo aéreo em relação ao ciclo terrestre (Quadro 8).

A taxa de substituição para a proteína L encontrada no ciclo terrestre do RABV foi aproximadamente 1,16 vezes em relação ao ciclo aéreo, porém esta variação não foi tão expressiva quanto à encontrada para os genes N, P e M.

Para inferir o tempo de divergência entre os ciclos aéreo e terrestre do RABV, refez-se a análise, desta vez utilizando os dois ciclos e apenas os genes G e L concatenados, que aparentemente estão evoluindo em taxas aproximadas (Quadro 8 e Figura 7).

A idade da diversificação dos ciclos do RABV em aéreo e terrestre a partir de um ancestral comum foi de 1870 anos, variando de 1178 a 2723 anos.

Na árvore filogenética de inferência bayesiana feita a partir dos genes G e L concatenados (Figura 7), o nó de origem entre os ciclos aéreo e terrestre apresentou a idade de 1783 anos, variando entre 1101 a 2937 anos.

Na figura 7 estão representadas todas as linhagens de RABV encontradas no ciclo aéreo e terrestre, de acordo com a classificação proposta por Bourhy et al., 2008 e Kuzmin et al., 2012 bem como as linhagens de RABV brasileiras sequenciadas neste estudo, *Myotis* Brasil (MY-BR), *Eptesicus* 1 Brasil (EP1-BR), *Eptesicus* 2 Brasil (EP2-BR), *D. rotundus* (DR), *Nyctinomops* Brasil (NY-BR), *T. brasiliensis* South America (TB-SA), *C. jacchus* (CJ-BR) e *C. thous* (CT-BR). As linhagens MY-BR, EP1-BR, EP2-BR, NY-BR e CJ-BR tiveram pela primeira vez todos os seus genes sequenciados.

Na figura 8 encontra-se a árvore de máxima verossimilhança inferida a partir das sequências genômicas nas quais os ramos de origem de cada agrupamento foram coloridos seguindo-se o padrão utilizado para a árvore de inferência bayesiana G-L. Embora a relação de ancestralidade para algumas linhagens do ciclo aéreo não serem as mesmas encontradas para a árvore de inferência bayesiana G-L, observa-se a recuperação das mesmas.

Quadro 8 – Taxas de substituição de nucleotídeos por sítio e TACMR encontrados para os genes N, P, M, G, L separadamente nos ciclos aéreos e terrestres e para os dois ciclos unidos utilizando-se os genes G+L concatenados - São Paulo – 2014

Gene	Taxa de Substituição		TMRCA	
	Aereo	Terrestre	Aereo	Terrestre
Nucleoproteína	1.30E-4 (1.22E-4 - 1.38 E-4)	2.26E-4 (1.40 E-4 - 3.14E-4)	998.524 (754.16 - 1266.86)	594.90 (370.228 - 876.97)
Proteína P	2.28 E-4 (2.04E-4 - 2.50E-4)	1.17E-4 (6.9E-5 - 1.64E-4)	710.11 (517.1 - 935.39)	1374.14 (839.82 - 2067.74)
Proteína M	2.49 E-4 (2.37E-4 -2.60E-4)	1.17 E-4 (6.79 E-5 - 1.62 E-4)	608.61 (450.84 - 780.42)	1100.99 (639.20 - 1679.46)
Proteína G	1.4E-4 (1.33E-4 - 1.47E-4)	1.18 E-4 (7.19E-5 - 1.65E-4)	1099.35 (807.90 - 1419.75)	1264.93 (771.61 - 1844.29)
Polimerase	1.15E-4 (1.13E-4 - 1.16E-4)	1.34E-4 (1.32E-4 - 1.36E-4)	1209.79 (974.65 - 1464.61)	1254.59 (947.11 - 1622.08)
G+L	1.14E-4 (7.09E-5 - 1.54 E-4)		1872.16 (1178 - 2723)	

Figura 7 – Árvores com máxima credibilidade de clados. A árvore foi inferida a partir dos genes G e L concatenados. Os ramos estão calibrados pelas datas de isolamento em anos. Na parte inferior está representada a escala dos eventos de separação entre as linhagens de RABV. Os grupos foram coloridos de acordo com a classificação proposta por Bourhy et al., 2008 e Kuzmin et al., 2012 - São Paulo - 2014

Figura 8 – Árvore filogenética inferida por máxima verossimilhança a partir de sequências genômicas. Os grupos foram coloridos de acordo com a classificação proposta por Bourhy et al., 2008 e Kuzmin et al., 2012 - São Paulo - 2014.

4.9 ANÁLISE DO SINAL FILOGENÉTICO

Como resultado desta análise, constatou-se que o gene G do RABV é o mais adequado para a realização de estudos filogenéticos (Quadro 9), apresentando o melhor resolução em relação a todos os outros quatro genes, Pseudo gene separadamente e sequência genômica.

Quadro 9 – Descrição do sinal filogenético encontrado para cada gene e genoma do RABV - São Paulo - 2014

Região genômica	Quartetos resolvidos (%)	Quartetos não resolvidos (%)*	Quartetos parcialmente resolvidos (%)
Nucleoproteína	95,9	1,7	2,4
Proteína P	94,9	2,5	2,6
Proteína M	94,0	2,7	3,3
Glicoproteína	99,6	0,1	0,3
Polimerase	98,9	0,3	0,8
Genoma	99,3	0,1	0,6

4.10 ANÁLISE DO REGIME DE SELEÇÃO

Como já descrito no item 3.18, a análise do regime de seleção atuante em cada gene do RABV foi realizada separadamente para os ciclos aéreo e terrestre.

Foi encontrado apenas um sítio sob seleção positiva, com $\omega = 5,49$ e $p = 0,03$ (Quadro 12), localizado no aminoácido 212 da proteína G do ciclo terrestre. Os aminoácidos e respectivos códons para este sítio estão descritos no quadro 10. Na proteína G madura (sem a região peptídeo-sinal) este aminoácido corresponde à posição 193 da glicoproteína.

Este aminoácido está localizado em um epítipo conformacional formado por duas regiões separadas, a região “neurotoxin-like” (aminoácidos 189-214) e o sítio antigênico III (aminoácidos 330-340), o qual participa da ativação da atividade fusogênica da glicoproteína do RABV, para a sua liberação no citoplasma celular (WUNNER, 2007).

Quadro 10 – Aminoácidos e respectivos códons para posição 193 da proteína G do ciclo terrestre - São Paulo - 2014

Aminoácido	Classe	Códons utilizados	Numero de amostras
Treonina (T)	Polar não carregado	ACC/ACT	43 (41 ACC)
Alanina (A)	Apolar	GCC	8
Valina (V)	Apolar	GTC	5
Isoleucina (I)	Apolar	ATC	3

Não foram encontradas relações entre a presença destes aminoácidos em grupos monofiléticos específicos da árvore filogenéticas com a máxima credibilidade de clados (MCC) do ciclo terrestre geradas a partir do gene G, demonstrando que tal substituição não é específica para apenas um nó terminal da árvore filogenética. Isto sugere que estas mutações podem ter ocorrido múltiplas vezes em diferentes linhagens do RABV.

Para todos os outros genes, nos dois ciclos foram encontradas somente evidências de seleção purificadora ($\omega < 0$) (Figura 9). Com exceção do gene L ($p = 0,03$) não foram encontradas diferenças estatisticamente significativas entre a quantidade de sítios sob seleção para os genes N, P, M e G nos ciclos terrestre e aéreo, demonstrando que os diferentes ciclos não são fatores determinantes para o acúmulo de sítios sob seleção nessas proteínas. No entanto, não se observa o mesmo padrão para o gene L, onde foram encontradas diferenças estatisticamente significativas entre a quantidade de sítios sob seleção purificadora entre os ciclos aéreo e terrestre, sendo que o ciclo aéreo apresentou mais sítios sob seleção, em relação ao ciclo terrestre (Quadro 11).

Quadro 11 – Número de sítios sob seleção purificadora nos 5 genes do RABV, porcentagem de sítios sob seleção purificadora em relação ao número de aminoácidos total de cada proteína e respectivos valores de p (teste exato de Fisher) - São Paulo - 2014

Ciclos	Gene N (450 aa)	%	Gene P (297 aa)	%	Gene M (202 aa)	%	Gene G (524 aa)	%	Gene L (2.127 aa)	%
Aereo	309	68,8	158	53,9	137	67,8	327	62,4	1529	71,8
Terrestre	290	64,6	161	54,9	116	57,4	321	61,2	1414	66,5
p-valor	0,44		0,87		0,19		0,81		0,03	

Quando analisados os valores de ω encontrados para cada gene nos ciclos aéreo e terrestre, encontrou-se diferenças estatisticamente significantes na intensidade da seleção purificadora atuante nos dois ciclos do RABV, mostrando que a seleção purificadora atua com maior intensidade no ciclo aéreo em relação ao ciclo terrestre (Quadro 12 e Figura 9).

Figura 9 – Intensidade da seleção purificadora ($\omega < 0$) atuante nos 5 genes do RABV e os respectivos valores de p - São Paulo - 2014

Quadro 12 – Valores de ω (dN-dS) para os cinco genes do RABV calculados separadamente para os ciclos aéreo e terrestre - São Paulo – 2014

Gene	Ciclo	Mediana	Mínimo	Maximo
N	Ar	-5,85	-28,73	-2,99
	Terra	-5,22	-14,69	-2,97
P	Ar	-6	-16,53	-3
	Terra	-5,15	-12,07	-2,99
M	Ar	-5,77	-15,14	-3
	Terra	-4,62	-12,46	-2,85
G	Ar	-6	-14	-2,97
	Terra	-5	-15,21	5,49
L	Ar	-5,9	-18	-2,78
	Terra	-5,15	-18,65	-2,76

vermelho: seleção positiva

4.11 ANÁLISE DO USO PREFERENCIAL DE CÓDONS

4.11.1 Índice de Adaptação de Códon (CAI)

No quadro 13 e figura 10, podemos observar que o gene M presente nas amostras do ciclo aéreo do RABV apresentou o maior valor de mediana (0,73), bem como os valores mínimo (0,685) e máximo (0,762) entre todos os outros genes do RABV em ambos os ciclos, sendo o único que não apresentou diferença estatisticamente significativa para os valores encontrados para o ciclo aéreo e terrestre.

Quadro 13 – Índice de Adaptação de Códon encontrado para cada gene do RABV nos ciclos aéreo e terrestre - São Paulo – 2014

Gene	Ciclo	Mediana	Mínimo	Maximo
N	Ar	0,714	0,692	0,729
	Terra	0,708	0,694	0,723
P	Ar	0,722	0,691	0,741
	Terra	0,715	0,699	0,74
M	Ar	0,73	0,685	0,762
	Terra	0,728	0,707	0,757
G	Ar	0,719	0,699	0,746
	Terra	0,708	0,694	0,72
L	Ar	0,705	0,698	0,714
	Terra	0,704	0,693	0,71

Figura 10 – Índice de Adaptação de Códon para os 5 genes do RABV analisados separadamente para os ciclos aéreo (azul) e terrestre (marrom) com os respectivos valores de p para a significância estatística das diferenças encontradas - São Paulo - 2014.

Azul – ciclo aéreo; Marrom – ciclo terrestre

Os RABV mantidos no ciclo aéreo apresentaram maiores valores de mediana para os 5 genes analisados em relação ao ciclo terrestre, sendo esta diferença estatisticamente significativa para os genes N, P, G e L, além de apresentarem os valores máximos de CAI para cada um dos 5 genes do RABV (Quadro 13 e figura 8).

Os resultados encontrados relativos ao índice de adaptação de códons mostram que os cinco genes dos RABV distribuídos nos ciclos aéreo e terrestre estão adaptados quanto sua expressão em seus reservatórios, quando comparados a um conjunto de códons de proteínas de alta expressão sabidamente bem adaptados em *Mus musculus*.

4.11.2 Número efetivo de códons utilizados (Nc) e conteúdo GC3.

No quadro 14 e figura 11 A estão expostos os resultados encontrados para o número efetivo de códons utilizados (Nc) nos cinco genes dos RABV circulantes no ciclo aéreo e

terrestre. Quanto à variação dos valores de mediana do Nc (52,9685 - 56,876), observamos pouco viés para utilização do número efetivo de códons sinônimos, caracterizados por valores de Nc abaixo de 40 (BELALOV; LUKASHEV, 2013).

Na figura 15B a 15F estão expostas as relações Nc vs GC3 encontradas para cada gene do RABV nos ciclos aéreo e terrestre.

Quadro 14 – Número efetivo de códons utilizados (Nc) para cada gene do RABV nos ciclos aéreo e terrestre - São Paulo – 2014

Gene	Ciclo	Mediana	Mínimo	Maximo
N	Ar	56,82	53,06	58,61
	Terra	56,27	53	60,02
P	Ar	58,69	49,53	61
	Terra	56,88	53,95	61
M	Ar	55,65	47,72	61
	Terra	53,35	49,15	58,45
G	Ar	52,97	50,64	57,18
	Terra	53,55	49,37	56,19
L	Ar	53,8	52,55	54,69
	Terra	53,32	52,55	54,55

Quadro 15 – Conteúdo GC3 encontrado para cada gene do RABV nos ciclos aéreo e terrestre - São Paulo - 2014

Gene	Ciclo	Mediana	Mínimo	Maximo
N	Ar	46,77	45,66	47,44
	Terra	48,11	47,22	48,99
P	Ar	54,61	51,54	58,02
	Terra	54,61	51,19	56,31
M	Ar	49,51	46,04	51,49
	Terra	49,01	46,54	51,49
G	Ar	49,62	47,9	51,72
	Terra	49,05	47,71	50,19
L	Ar	45,28	44,57	45,84
	Terra	45,28	43,96	45,93

Figura 11 – A) boxplot mostrando o Nc de cada gene do RABV separado em ciclo aéreo (azul) e ciclo terrestre (marrom) com os respectivos valores de p para a significância estatística para as diferenças encontradas entre os dois ciclos; B a F) cinco gráficos mostrando o esperado (curva no gráfico) e o observado (pontos no gráfico) efetivo número de códons (Enc e Nc respectivamente) em relação ao conteúdo GC3 observado para os genes N, P, M, G e L do RABV nos ciclos aéreo (azul) e terrestre (marrom) do RABV - São Paulo - 2014

O menor valor absoluto para o Nc foi encontrado no gene M do ciclo aéreo (47,718), enquanto os maiores valores absolutos de Nc (61) foram encontrados para o gene M do ciclo aéreo e gene P de ambos os ciclos.

4.11.3 Uso relativo de códons sinônimos (RSCU)

Quanto ao uso relativo de códons sinônimos observa-se que diversos aminoácidos em ambos os ciclos do RABV utilizam preferencialmente alguns códons sinônimos para a sua tradução, sendo que a proporção de utilização da maioria dos códons

sinônimos para determinados aminoácidos é diferente entre os ciclos aéreo e terrestre do RABV ($p < 0,05$)

Quadro 16 – Distribuição por amostra dos códons utilizados preferencialmente em cada gene do RABV nos ciclos aéreo e terrestre - São Paulo - 2014

		NÚMERO DE AMOSTRAS QUE UTILIZAM PREFERENCIALMENTE ALGUM CODON EM CADA CICLO														
		GENES														
		N			P			M			G			L		
	CODON	AR	TERRA	p	AR	TERRA	p	AR	TERRA	p	AR	TERRA	p	AR	TERRA	p
Arginina (Arg)	AGA	100	59	2,20E-03	99	59	9,51E-05	100	59	7,58E-01	100	59	1,42E-05	100	59	1
	AGG	97	49		89	53		100	59		97	28		100	59	
	CGU	0	0		16	1		0	0		0	0		0	0	
	CGC	0	0		0	0		0	0		0	0		0	0	
	CGA	0	7		26	32		1	0		0	7		0	0	
	CGG	1	1		9	18		3	0		0	0		0	0	
Histidina (His)	CAU	91	30	1,80E-06	85	36	7,56E-01	61	39	3,14E-01	98	32	9,49E-09	98	50	3,84E-03
	CAC	8	22		10	3		9	10		2	19		1	7	
Lisina (Lys)	AAA	2	15	6,26E-06	67	10	4,48E-01	15	28	2,02E-03	13	21	1,46E-04	48	4	6,97E-09
	AAG	95	42		25	46		45	24		86	29		46	53	
Acido aspartico (Asp)	GAU	10	43	< 2,2E-16	5	45	< 2,2E-16	98	23	3,84E-11	45	24	1	88	14	< 2,2E-16
	GAC	87	9		89	14		2	21		48	27		9	42	
Acido glutâmico (Glu)	GAA	3	3	6,61E-01	0	7	7,65E-04	18	7	2,62E-01	8	0	2,37E-02	0	2	1,36E-01
	GAG	93	50		100	52		70	51		85	58		100	57	
Serina (Ser)	UCU	100	52	8,87E-07	98	50	4,51E-05	100	59	< 2,2E-16	99	43	5,53E-06	100	59	1
	UCC	8	32		94	41		11	14		41	52		97	58	
	UCA	98	57		61	58		92	27		100	57		100	59	
	UCG	2	4		0	7		2	11		0	3		0	0	
	AGU	30	30		4	0		0	59		4	12		0	0	
	AGC	0	0		6	8		0	3		0	0		0	0	
Treonina (Thr)	ACU	88	57	1,23E-02	90	50	4,61E-01	70	47	7,31E-03	20	20	1,03E-01	92	55	1
	ACC	76	24		79	55		28	27		99	59		99	58	
	ACA	93	34		7	3		74	30		62	27		100	59	
	ACG	0	1		0	1		0	3		0	0		0	0	
Asparagina (Asn)	AAU	90	41	5,52E-02	14	42	3,52E-14	2	4	1,97E-01	20	15	4,23E-01	4	3	7,11E-01
	AAC	7	9		82	13		97	55		67	36		96	56	
Glutamina (Gln)	CAA	23	19	1,31E-03	12	6	1	80	45	2,25E-01	88	28	1,28E-08	20	15	3,15E-01
	CAG	67	14		82	44		17	15		6	25		75	37	
Tirosina (Tyr)	UAU	93	59	1,23E-02	14	48	4,23E-06	87	19	7,10E-13	7	3	7,44E-01	86	58	1,52E-01
	UAC	7	0		21	7		12	40		90	55		7	1	
Leucina (Leu)	UUA	8	20	1,65E-03	3	13	4,27E-10	6	2	6,78E-07	0	0	4,43E-03	2	0	1,81E-10
	UUG	100	52		56	31		99	46		84	54		100	59	
	CUU	5	10		2	18		60	51		90	57		54	11	
	CUC	1	0		93	52		9	1		21	18		13	1	
	CUA	49	34		36	5		13	39		34	4		0	7	
	CUG	71	56		100	44		58	44		95	51		99	55	
Valina (Val)	GUU	78	41	7,57E-11	24	21	5,07E-03	27	17	3,93E-01	35	56	1,64E-09	11	27	1,21E-05
	GUC	32	54		89	57		83	47		79	17		96	37	
	GUA	5	2		0	5		0	2		0	0		2	0	
	GUG	84	14		37	13		76	44		94	45		97	56	
Alanina (Ala)	GCU	31	52	6,66E-05	95	58	2,31E-01	93	56	5,00E-06	83	12	2,96E-06	16	50	7,35E-08
	GCC	63	31		94	55		9	21		5	8		95	57	
	GCA	100	58		6	0		86	47		100	59		100	59	
	GCG	0	0		1	0		18	0		0	0		0	0	
Prolina (Pro)	CCU	93	53	6,98E-15	47	59	4,45E-10	92	53	8,15E-06	99	53	6,36E-05	100	59	1,92E-02
	CCC	7	18		87	27		57	12		61	35		80	50	
	CCA	8	11		65	21		85	10		17	34		1	7	
	CCG	62	0		1	11		9	6		0	3		0	0	
Glicina (Gly)	GGU	5	19	3,09E-04	10	15	3,87E-07	0	2	1,84E-04	14	5	6,71E-01	0	0	1
	GGC	2	1		9	1		81	41		0	0		0	0	
	GGA	99	57		85	25		89	51		88	48		100	59	
	GGG	81	43		95	16		15	30		99	58		100	59	
Isoleucina (Ile)	AUU	35	34	2,11E-02	25	2	1,72E-05	80	7	2,90E-08	0	11	1,19E-06	0	0	3,83E-01
	AUC	9	14		30	39		90	59		96	31		97	59	
	AUA	95	51		96	59		4	5		90	49		58	30	
Cisteina (Cys)	UGU	26	33	3,86E-01	100	58	1	81	56	1,57E-01	85	49	4,80E-01	93	42	1,41E-03
	UGC	9	6		0	0		5	0		12	10		4	12	
Fenilalanina (Phe)	UUU	23	1	1,38E-04	14	13	1,90E-01	76	22	3,13E-08	12	24	6,36E-05	49	28	1
	UUC	77	58		73	38		5	22		81	29		47	26	

4.12 DIVERSIDADE GENÉTICA CALCULADA ATRAVÉS DA DISTÂNCIA PATRÍSTICA E SUA COMPARAÇÃO COM OS RESULTADO DE DISTÂNCIA GENÉTICA E IDENTIDADE GENÉTICA.

4.12.1 Análise em separado dos ciclos aéreo e terrestre do RABV

Na figura 19 e quadro 16 observa-se que os maiores valores de mediana de distância patrística para todos os genes do RABV encontram-se no ciclo terrestre em relação ao ciclo aéreo, sendo que as diferenças dos valores de mediana entre as distâncias patrísticas para os mesmos genes nos dois ciclos foram estatisticamente significativas, com valores de $p < 2.2E-16$. Isto pode ser interpretado como um maior número de substituições de nucleotídeos por sítio nos 5 genes do RABV ocorrendo entre a maioria das amostras de RABV circulantes no ciclo terrestre em relação ao ciclo aéreo.

Figura 12 – Estimativa da distância patrística para os genes N, P, M, G e L encontradas nos ciclos aéreo e terrestre do RABV- São Paulo - 2014

Quadro 17 – Valores medianos, mínimos e máximos encontrados para os valores de distância patrística nos ciclos aéreo e terrestre do RABV analisados separadamente e em conjunto Paulo - 2014

Gene	Ciclos	Mediana	Mínimo	Maximo
N	Aéreo	0,13	2,00E-08	0,38
	Terrestre	0,20	2,00E-08	0,30
	Aéreo e Terrestre	0,25	2,00E-08	0,48
P	Aéreo	0,18	2,00E-08	0,54
	Terrestre	0,28	2,00E-08	0,39
	Aéreo e Terrestre	0,32	2,00E-08	0,60
M	Aéreo	0,11	2,00E-08	0,32
	Terrestre	0,18	2,00E-08	0,32
	Aéreo e Terrestre	0,26	2,00E-08	0,48
G	Aéreo	0,16	2,00E-08	0,49
	Terrestre	0,25	2,00E-08	0,37
	Aéreo e Terrestre	0,28	2,00E-08	0,51
L	Aéreo	0,11	2,00E-08	0,33
	Terrestre	0,22	0,000276	0,30
	Aéreo e Terrestre	0,27	2,00E-08	0,51

De acordo com os valores de mediana das distâncias patrísticas (Quadro 16), podemos observar o seguinte grau de conservação genética para cada gene em cada um dos ciclos: ciclo aéreo $M > L > N > G > P$; ciclo terrestre $M > N > L > G > P$.

Quadro 18 – Variabilidade genética encontrada para os 5 genes do RABV calculada através dos métodos de distância patrística, identidade genética e distância genética, analisadas para os ciclos aéreo e terrestre do RABV- São Paulo - 2014

Gene	Ciclos	Distância patrística	Identidade genética	1- Distância genética
N	Áereo e terrestre	0,25	0,86	0,88
	Áereo	0,13	0,92	0,91
	Terrestre	0,20	0,89	0,88
P	Áereo e terrestre	0,32	0,85	0,82
	Áereo	0,18	0,89	0,88
	Terrestre	0,28	0,86	0,84
M	Áereo e terrestre	0,26	0,87	0,85
	Áereo	0,11	0,91	0,90
	Terrestre	0,18	0,89	0,88
G	Áereo e terrestre	0,28	0,85	0,83
	Áereo	0,16	0,89	0,88
	Terrestre	0,25	0,87	0,85
L	Áereo e terrestre	0,27	0,87	0,90
	Áereo	0,11	0,91	0,94
	Terrestre	0,22	0,88	0,86

Os resultados da variabilidade genética para cada gene encontrada em cada ciclo do RABV obtidos através de uma matriz de identidade genética foram os seguintes: ciclo aéreo $N < M < L < P < G$; ciclo terrestre $N < M < L < G < P$ (Quadro 16).

Quando analisamos a variabilidade genética a partir do cálculo de distância genética (Quadro 16) utilizando-se o modelo evolutivo MCL observamos os seguintes resultados: ciclo aéreo $L < N < M < G < P$; ciclo terrestre $N = M < L < G < P$.

As comparações entre as divergências genéticas estimadas para cada gene nos ciclos aéreo e terrestre do RABV através da distância patrística, distância genética e identidade genética encontram-se no quadro 16.

Os genes mais conservados e mais variáveis de acordo com o modelo utilizado e o ciclo do RABV encontram-se nos quadros 18 e 19.

Quadro 19 – Genes mais conservados nos ciclos aéreo e terrestre do RABV, analisados separadamente e em conjunto, comparando os resultados de distância patrística, distância genética e identidade genética - São Paulo - 2014

Ciclos	Distância patrística	Distância genética	Identidade genética
Aéreo	M = 0,11	L = 0,94	N = 0,92
Terrestre	M = 0,18	N e M = 0,88	M = 0,89
Aéreo e Terrestre	N = 0,25	L = 0,90	N = 0,88

Quadro 20 – Genes mais variáveis nos ciclos aéreo e terrestre do RABV, analisados separadamente e em conjunto, comparando os resultados de distância patrística, distância genética e identidade genética - São Paulo - 2014

Ciclos	Distância patrística	Distância genética	Identidade genética
Aéreo	P = 0,18	G = ,088	G = 0,89
Terrestre	P = 0,28	P = 0,84	P = 0,86
Aéreo e Terrestre	P = 0,32	P = 0,82	P = 0,85

Os coeficientes de correlação entre distância patrística distância genética e identidade genética encontrados para os cinco genes do RABV nos ciclos aéreo e terrestre são mostrados nos quadros 21 e 22 respectivamente.

Quadro 21 – Coeficiente de correlação para as divergências genéticas encontradas para os cinco genes do RABV no ciclo aéreo através dos métodos de distância patrística, distância genética e identidade genética - São Paulo - 2014

Ciclo aéreo	Distância patrística	Distância genética	Identidade genética
Distância patrística	#	#	#
Distância genética	0,83	#	#
Identidade genética	0,89	0,77	#

Quadro 22 – Coeficiente de correlação para as divergências genéticas encontradas para os cinco genes do RABV no ciclo terrestre através dos métodos de distância patrística, distância genética e identidade genética - São Paulo - 2014

Ciclo terrestre	Distância patrística	Distância genética	Identidade genética
Distância patrística	#		
Distância genética	0,97	#	
Identidade genética	0,98	0,99	#

4.12.2 Análise em conjunto dos ciclos aéreo e terrestre do RABV

Na figura 13 e quadro 17 estão descritas as distâncias patrísticas encontradas para os cinco genes do RABV analisando os ciclos aéreo e terrestre conjuntamente, sendo que diferenças entre as distâncias patrísticas para cada gene foram estatisticamente significativas (p value $< 2.2e-16$, Kruskal-Wallis rank sum test, $\alpha = 0.05$).

De acordo com os valores de mediana das distâncias patrísticas (Quadro 18 e Figura 17), podemos observar que o seguinte grau de variabilidade genética para cada gene do RABV : $N < M < L < G < P$.

Os resultados da variabilidade genética do RABV obtidos através da distância patrística, distância genética e identidade genética estão no quadro 18.

Os resultados da variabilidade genética do RABV obtidos à partir do cálculo da identidade genética para cada gene encontrada para o RABV foram os seguintes: $N < M < L < P < G$.

Já quando analisamos a variabilidade genética do RABV usando a distância genética obtivemos os seguintes resultados: $L < N < M < G < P$.

Os coeficiente de correlação entre distância patrística, distância genética e identidade genética encontrados para os cinco genes do RABV são mostrados no quadro 23.

Quadro 23 – Coeficiente de correlação para as divergências genéticas encontradas para os cinco genes do RABV através dos métodos de distância patrística, distância genética e identidade genética - São Paulo - 2014

RABV	Distância patrística	Distância genética	Identidade genética
Distância patrística	#	#	#
Distância genética	$r^2 = 0,67$	#	#
Identidade genética	$r^2 = 0,94$	$r^2 = 0,71$	#

Figura 13 – Distâncias patrísticas encontradas para os cinco genes do RABV analisando os ciclos aéreo e terrestre conjuntamente - São Paulo - 2014

5 DISCUSSÃO

Com o objetivo de estudar as relações filogenéticas e evolução dos RABV por meio da análise dos genomas de diversas linhagens mundialmente distribuídas, esta tese de doutorado foi dividida em duas etapas. Primeiramente, padronizou-se uma técnica de RT-PCR com o objetivo de amplificar em uma única reação o genoma de oito linhagens de RABV encontradas no Brasil para o seu posterior sequenciamento e montagem como descrito na seção Material e Métodos.

Após a conclusão desta primeira etapa, iniciou-se a segunda etapa deste projeto de tese de doutorado, que consistiu na busca e escolha de sequências de genomas de RABV na base de dados GenBank, inclusão neste conjunto de dados das sequências geradas neste estudo, para então realizar as análises filogenéticas e evolutivas propostas nesta pesquisa.

5.1 PADRONIZAÇÃO DA REAÇÃO DE RT-PCR GENÔMICA DO RABV E SUA UTILIZAÇÃO PARA A REALIZAÇÃO DO SEQUENCIAMENTO DE DNA POR AMPLIFICAÇÃO NA PLATAFORMA ILLUMINA MISEQ

A técnica de RT-PCR para a amplificação do genoma completo de diferentes linhagens do RABV a partir da extremidade 3' do RNA anti genômico, mostrou-se viável e eficaz, como demonstrado pelo sucesso em se obter o amplicon de tamanho genômico para 7 linhagens do RABV relacionadas a hospedeiros de espécies diversas, sendo que o isolado 3208/06, representante da linhagem *Eptesicus II*, foi amplificado por meio das reações de PCR A e B.

Entre estas oito linhagens, cinco ainda não tinham todos os seus genes e regiões intergênicas G-L sequenciadas (*Eptesicus I*, *Eptesicus II*, *Nyctinomops*, *Myotis I*, e *C. jacchus*).

Com a realização de apenas uma etapa de amplificação, problemas como a diminuição da sensibilidade da técnica devido ao excesso de manipulação do c-DNA são

minimizados. Outra grande vantagem de se trabalhar com um único amplificado é a confiabilidade gerada na reação de sequenciamento de DNA, uma vez que sequenciando diferentes amplicons, mesmo que relativos à mesma amostra viral, pode-se eventualmente gerar amplicons de diferentes subpopulações virais para cada um dos segmentos em estudo em função de diversidade de hibridação dos *primers* em uso.

Embora o sequenciamento de genomas completos de diferentes isolados de RABV venha sendo realizado desde 1986 (TORDO et al., 1986), o sequenciamento a partir de uma única RT-PCR ainda não foi descrito Campos et al. (2011) padronizaram um protocolo para a transcrição reversa do RNA anti genômico da linhagem *D. rotundus* do RABV e sua amplificação, porém, esta PCR genômica não foi utilizada diretamente para o sequenciamento deste isolado, realizado a partir de diferentes reações de PCR do c-DNA genômico.

Ainda, a presente pesquisa foi pioneira, pois utilizou pela primeira vez no Brasil a tecnologia de sequenciamento maciço (*Deep sequencing*) em amostras de RABV, através da técnica de sequenciamento por síntese na plataforma Illumina Miseq. Até o momento, existe apenas uma publicação utilizando esta tecnologia para o sequenciamento do RABV (BORUCKI et al., 2013).

Para confirmar a transcrição reversa do RNA anti genômico a partir da região 3' UTR, para a sua utilização para a amplificação dos genomas completos do RABV, na presente pesquisa utilizou-se uma reação de PCR (PCR controle), visando amplificar 1533 pb a partir da extremidade 5' UTR do c-DNA obtido na transcrição reversa do RNA anti genômico, indicando a obtenção de tal amplicon que a síntese de c-DNA de tamanho genômico havia sido obtida, o que permitiu aferir-se a eficiência da transcrição reversa antes da amplificação genômica para o sequenciamento.

As etapas de aferição da eficiência das reações de síntese e amplificação de c-DNA anti genômico através da PCR controle mostraram-se úteis no que se refere à economia de recursos, tempo e materiais de consumo, pois serviram de triagem para a determinação das amostras ideais para a realização da PCR genômica para posterior sequenciamento de DNA.

A aferição da amplificação do c-DNA anti genômico dos RABV utilizados nesta pesquisa foi realizada com sequenciamento de DNA através do método de Sanger, com os mesmos primers utilizados para a amplificação dos c-DNAs dirigidos às extremidades 3' (primer Início) e 5' (primer Final) do RABV.

Uma vez que DNAs polimerases desprovidas de atividade corretiva exonuclease 3'-5' podem levar a inserção de “falsas” mutações em uma sequência de DNA (KISSI et al., 1999), foi utilizado na presente pesquisa um protocolo de PCR genômica com enzima DNA

polimerase com atividade corretiva exonuclease 3'-5 (GoTaq® Long PCR Master Mix), com o intuito de se elevar a especificidade na detecção de mutações por maior confiabilidade em cada posição sequenciada em relação aos nucleotídeos realmente presentes nos genomas virais em estudo. Isto é um ponto fundamental a ser considerado em estudos baseados em sequências de nucleotídeos, tão importante quanto à acurácia dos procedimentos bioinformáticos que se seguem a obtenção das mesmas.

A utilização do *Deep sequencing* apresenta diversas vantagens em relação aos sequenciamentos de DNA realizados pelo método de Sanger, entre elas: a) maior confiabilidade dos resultados obtidos, uma vez que fornece grande acurácia em cada posição dos nucleotídeos devido ao grande número de leituras de cada posição; b) como o sequenciamento por amplificação não utiliza *primers* específicos para o RABV, não existe a probabilidade de polimorfismos de posições devido à hibridação inespecífica de *primers*; c) menor tempo para a obtenção das sequências, uma vez que todos os amplicons são sequenciados em uma única reação, enquanto que pelo método de Sanger seriam necessários por volta de 30 reações de sequenciamento com *primers* específicos (15 reações com *primers* senso e 15 reações com *primers* anti-senso) para amplificar fragmentos de aproximadamente 800 pb, com apenas uma leitura de cada nucleotídeo, como proposto por Mochizuki et al., 2011; d) para cada amplicon sequenciado, referente a cada amostra viral, foi utilizado apenas 1 ng de DNA, enquanto que para o sequenciamento pelo método de Sanger, onde seriam necessários de 5 a 20 ng de DNA para sequenciar cada fragmento de 800 pb, segundo as instruções do fabricante.

5.2 ANÁLISE DESCRITIVA DAS LINHAGENS ESTUDADAS

As inserções ou deleções intraespecíficas encontradas em regiões não codificantes das amostras estudadas nesta pesquisa são comuns em linhagens do RABV. Isto é facilmente observado ao analisar o alinhamento completo utilizado na presente pesquisa (159 sequências com 11.785 posições). Estes fenômenos ocorrem por erros gerados pela RNA polimerase RNA dependente viral durante a replicação viral, ocorrendo geralmente em locais dos genomas sob baixa pressão de seleção sendo assim fixados (DUFFY et al., 2008), sendo que o

grande número de leituras de cada posição dos nucleotídeos obtidos através do sequenciamento por síntese na plataforma Illumina Miseq nos permite assegurar a existência destas inserções ou deleções.

A qualidade de nosso sequenciamento foi também investigada também através da comparação da amostra fixa CVS-11 mantida no IP com a amostra CVS-11 mantida pelo National Institute for the Control of Pharmaceutical and Biological Products da China. Nesta análise foram encontradas quatro diferentes fixações de nucleotídeos, duas no gene G e duas no gene L, sendo todas transições, três não sinônimas e uma sinônima. (Quadro 7).

Especulações sobre o que levaram a estas substituições seriam mais facilmente discutidas se fossem conhecidos alguns aspectos do método de manutenção da amostra CVS-11 chinesa, principalmente o tipo de sistema biológico utilizado e número de passagens da amostra em questão, mas esta informação não está disponível. No que diz respeito às três transições não sinônimas, as substituições de aminoácidos resultantes da substituição de nucleotídeos representam mudanças fenotípicas ocorridas entre as duas amostras de CVS em questão, observando-se que as substituições não sinônimas ocorreram apenas entre aminoácidos hidrofílicos, o que é biologicamente mais provável do ponto de vista funcional e estrutural em relação à proteína traduzida (AGRANOVICH; MAMAN; LOUZOUN, 2013).

A variação do aminoácido 408 da glicoproteína viral (que corresponde à posição 389 da glicoproteína madura), representados pelos aminoácidos Lisina (K) na amostra CVS-11 GQ918139 e Ácido glutâmico (E) na amostra CVS-11 mantida no IP, ocorreu entre dois aminoácidos hidrofílicos (polares) de características físico químicas diferentes, já que a Lisina é polar neutro, enquanto o Ácido glutâmico é polar negativo. Este aminoácido está localizado no ectodomínio da glicoproteína viral (aminoácidos 1 a 439), e não está em nenhum dos sítios antigênicos já mapeados para esta região (WUNNER, 2007), o que permite especular que esta mutação não inviabiliza a manutenção da amostra CVS-11 em laboratório.

A variação do aminoácido 517 da glicoproteína (aminoácido 498 da glicoproteína madura), entre os estados Arginina ↔ Lisina (R ↔ K), é viável biologicamente, uma vez que é uma variação de aminoácidos similares, ou seja, da mesma classe físico-química (polar neutro), o que geralmente não prejudicaria a conformação espacial e atividade biológica da proteína traduzida. É fundamental atentar para o fato de que este aminoácido está localizado na região da glicoproteína conhecida como domínio citoplasmático (aminoácidos 462 a 524), onde interage com a proteína M auxiliando na formação da estrutura e montagem do RABV, pois esta substituição entre aminoácidos de mesma classe físico-química traz um custo menor a função estrutural desta região de G.

A mudança do aminoácido 1588 da proteína L Asparagina $\leftrightarrow$ Ácido aspártico (N $\leftrightarrow$ D), embora não tenha ocorrido entre aminoácidos da mesma classe físico-química, uma vez que L é polar neutro e D é polar negativo, parece não ter tido consequências em redução de adaptabilidade (fitness) para estas duas amostras de CVS, uma vez que elas estavam se replicando nos sistemas dos quais elas foram isoladas.

Dado que a região em que se observou esta variação não faz parte dos sítios biologicamente ativos já descritos para a RNA polimerase do RABV, talvez exista pouca restrição para a manutenção da carga elétrica deste aminoácido.

Entretanto, a consequência biológica de todas as mutações encontradas nesta pesquisa deveriam ser acessadas de modo mais consistente por ensaios dirigidos a quantificação mais acurada da transcrição e replicação viral como qPCR para mensuração de fitness de modo comparado entre as duas amostras de CVS em questão.

Não foi possível estabelecer quais as substituições de nucleotídeos ocorreram no códon 450 da nucleoproteína da amostra de RABV 1770/12 (linhagem *C. jacchus*), responsáveis pela mudança do aminoácido esperado para esta posição em um códon de parada TGA. Nos outros dois isolados pertencentes a esta linhagem, foram encontrados os aminoácidos isoleucina (Q), traduzido pelo códon CAA, e o aminoácido S (serina), traduzido pelo códon TCA.

Como descrito por Finke et al. (2000), observou-se nas amostras sequenciadas nesta pesquisas, bem como nas outras sequências de RABV utilizadas no alinhamento completo deste trabalho, o seguinte tamanho das 4 regiões intergênicas do RABV: N/P 2 nt, P/M 5 nt, M/G 5 nt e G/L 24 nt.

Sendo assim não foi encontrada nas amostras de RABV estudadas aqui a região intergênica G/L de 423 nt, conhecida como pseudo gene, o que era esperado, pois sabe-se que esta região é uma exceção à regra, sendo encontrada excepcionalmente em algumas amostras fixas do RABV como PV, ERA e SAD1-3670, e não descrita para nenhum outro *Lyssavirus* (WUNNER, 2007).

A posição 333 na glicoproteína G madura, conhecida como arginina 333 (R333), por este ser o aminoácido mais comumente encontrado nesta posição na maioria das amostras selvagens de RABV, é descrito como importante para a neuroinvasividade e virulência associados aos RABV. Além da R, o aminoácido lisina (K) também já foi descrito para esta posição em amostras selvagens de RABV, enquanto substituições desta posição por glutamina (Q), isoleucina (I), glicina, metionina (M) ou serina (S) já foram descritas em vírus fixos ou

passagens de vírus selvagens em diferentes sistemas biológicos (SATO et al., 2004; WUNNER, 2007).

No alinhamento completo aqui utilizado (159 sequências), o qual utilizou apenas sequências selvagens, na posição 333 da glicoproteína madura do RABV, o aminoácido encontrado para a maioria das amostras foi a arginina (145 amostras), seguidos de K (5 amostras), Q (4 amostras), N (3 amostras) e H (2 amostras), sendo que a fixação do aminoácido N nesta posição foi exclusiva para os RABV da linhagem *E. furinalis* I enquanto a substituição K foi exclusiva para o complexo de linhagens mantidas por diferentes espécies do gênero *Lasiurus* nos EUA.

Entre todas estas fixações de aminoácidos para a posição 333 da glicoproteína do RABV, os aminoácidos N e H ainda não haviam sido descritos, e seu achado torna possível o delineamento de futuros estudos comparativos com o objetivo de estudar a virulência e adaptabilidade de RABV com diferentes aminoácidos para esta posição mantidos em diferentes sistemas biológicos.

5.3 ANÁLISES DE RECOMBINAÇÃO

A recombinação entre diferentes linhagens do RABV é um evento raro, e até o momento, foram descritos na literatura apenas dois eventos desta natureza, em linhagens de RABV pertencentes ao ciclo terrestre e mantidas por cães na China (LIU et al., 2011; HE et al., 2012). Para que aconteça a recombinação entre linhagens do RABV, é necessário que exista um hospedeiro infectado simultaneamente (coinfecção) com no mínimo duas linhagens distintas de RABV, para que estas se recombinem. Embora pouco provável, a recombinação entre diferentes vírus da raiva é possível, sendo que quando presente causam erros em qualquer tipo de análise que venha a utilizar as regiões sob recombinação (CHARE, E. R. , GOULD, E. A., HOLMES, 2003; HAN; WOROBEY, 2011).

A coinfeção de células hospedeiras por diferentes linhagens virais ocorre com baixa frequência em vírus de RNA de sentido negativo devido ao fato de que estes vírus geralmente produzem infecções agudas com curtos períodos até a recuperação ou desfecho letal, acarretando em uma pequena probabilidade de múltiplas infecções em um hospedeiro o que

limita a detecção da recombinação. As melhores evidências de recombinações na ordem *Mononegavirales* foram encontradas em arnavírus e hantavírus, os quais são responsáveis por infecções persistentes em roedores (CHARE; GOULD; HOLMES, 2003; HAN; WOROBEY, 2011).

Outra possível razão para a baixa taxa de recombinação na ordem *Mononegavirales* é o próprio complexo ribonucleocapsídeo, o qual nunca está dissociado do RNA viral dificultando a mudança da RNA polimerase viral de um genoma para o outro durante o processo de replicação. Por outro lado, a produção de RNAs defectivos em vírus RNA negativos mostra que a RNA polimerase viral pode saltar em uma mesma molécula ou para outras moléculas de RNA com relativa frequência. Desta maneira, o complexo RNP pode não representar uma séria barreira para a recombinação (CHARE, E. R. , GOULD, E. A., HOLMES, 2003; HAN; WOROBEY, 2011).

5.4 ANÁLISE DE HETEROTAQUIA

A hipótese da existência de heterotaquia entre os ciclo aéreo e terrestre do RABV foi testada e comprovada uma vez que o menor valor de verossimilhança foi obtido na hipótese em que as linhagens pertencentes aos ciclos aéreo e terrestre da raiva estão sob relógios moleculares distintos, com diferença estatisticamente significativa em relação ao valor de verossimilhança encontrado para a hipótese em que ambos os ciclos estariam sob um mesmo relógio molecular.

A segregação dos RABV nos ciclos aéreo (mantido principalmente por reservatórios da ordem *Chiroptera*) e terrestre (mantido exclusivamente por reservatórios da ordem *Carnivora*) já é bem conhecida e aceita-se a hipótese de que as características inerentes aos reservatórios existentes em cada um destes ciclos levaram à sua segregação nestes dois grandes grupos (BOURHY et al., 2008; KUZMIN et al., 2012).

Até o presente momento, não há relato da existência da heterotaquia entre os mesmos genes dos RABV mantidos nos ciclos aéreo e terrestre, exceto aquela entre cada gene em um mesmo genoma (WU et al., 2007). Desta maneira, esta é a primeira descrição deste fenômeno para o RABV em um nível populacional, e propõe-se que sua existência também esteja

relacionada às características inerentes as diferentes ordens de reservatórios responsáveis pela manutenção destes dois ciclos.

Como já citado, heterotaquia é o fenômeno no qual as taxas de evolução de determinados sítios ou regiões codificantes de determinado gene podem variar em uma dada filogenia. Estas diferentes taxas de evolução podem ser relacionadas a processos seletivos distintos atuando nas sequências sob heterotaquia (LOCKHART et al., 2006). Diante disto, pode-se inferir que determinados processos seletivos para a adaptação dos RABV em cada ciclo podem estar atuando de maneiras diferentes em um mesmo gene.

Embora tais análises sejam realizadas gene a gene, ou proteína a proteína, não podemos esquecer que a interação RABV- célula hospedeira representa um sistema biológico complexo e adaptativo, que tem como propriedade emergente a partícula viral que só pode ser explicada pela interação de todos os componentes deste sistema e não por cada um deles individualmente.

Desta maneira o RABV interage como um todo com a célula hospedeira e a adaptação e evolução do vírus aos diferentes reservatórios só pode ser explicada pela interação de todos os componentes do sistema (sistema imune, proteínas virais, proteínas celulares, organelas celulares, RNAs, etc). As proteínas do RABV exercem interações exclusivas entre si e com diferentes tipos de moléculas da célula hospedeira para que ocorram as interações necessárias para a sua entrada e replicação nas células do hospedeiro (SCHNELL et al., 2010).

A estratégia do RABV em evadir-se do sistema imune adaptativo também tem grande contribuição na evolução viral (SCHNELL et al., 2010), sendo que em cada ciclo, esta interação deve ser hospedeiro-específica. Em contraste com a raiva praticamente 100% letal em carnívoros e outros mamíferos, estudos realizados em colônias de morcegos hematófagos *D. rotundus* sugerem que nesta espécie a probabilidade de desenvolver uma infecção letal parece ser bastante baixa, em torno de apenas 10% dos animais infectados. A recuperação após a infecção também pode ser demonstrada pela grande quantidade de morcegos naturalmente imunizados encontrados na natureza e em estudo de inoculação experimental (STEECE; ALTENBACH, 1989; JACKSON et al., 2008; BLACKWOOD et al., 2013).

A proteína P exerce papel fundamental no mecanismo de escape do sistema imune inato da célula hospedeira, interferindo de maneira ativa na produção de interferons 1 e 2. A proteína G tem papel fundamental na ausência de apoptose celular relacionada aos RABV patogênicos e ao curso letal da doença (SCHNELL et al., 2010).

Diante do exposto, é sugestivo que diferentes nos processos seletivos, ou diferentes intensidade destes, estão atuando nos 5 genes do RABV de acordo com o ciclo analisado,

relacionados provavelmente a diferentes interações vírus-hospedeiro ocorrendo em cada ciclo, o que em última análise se reflete nos diferentes relógios moleculares a que estão submetidos os ciclos aéreo e terrestre do RABV.

Para estudar mais detalhadamente o fenômeno da heterotaquia existente entre os ciclos aéreo e terrestre do RABV decidiu-se dividir parte da análise realizada nesta pesquisa, analisando separadamente cada ciclo. Desta maneira torna-se possível estabelecer relações entre a heterotaquia e as diferenças estatisticamente significativas dos resultados encontrados nas análises realizadas separadamente para cada ciclo.

Como será discutido nos itens seguintes, após obtidos os resultados dos estudos evolutivos e filogenéticos realizados separadamente para cada gene em cada ciclo do RABV, foi possível estabelecer quais seriam os melhores genes para utilizar na realização de análises filogenéticas confiáveis do RABV como um todo.

5.5 ESTIMATIVAS DA TAXA DE SUBSTITUIÇÃO DE NUCLEOTÍDEOS POR SÍTIO POR ANO E TEMPO DE ANCESTRAL COMUM MAIS RECENTE (TACMR) E ESTIMATIVA DE FILOGENIAS

Neste item, discutiremos as diferentes taxas de substituição de nucleotídeos (heterotaquia) intraciclo, isto é, entre todos os 5 genes do RABV em um mesmo ciclo, e a heterotaquia interciclos, ou seja, as diferenças encontradas entre os mesmos genes em cada ciclo.

No ciclo terrestre, as maiores taxas de substituição foram em média encontradas para o gene N ($2.26E-4$), as quais foram aproximadamente o dobro das taxas encontradas para os outros genes, que parecem evoluir em taxas aproximadas, sendo que as taxas de substituição de nucleotídeos para os genes P ($1.17E-4$), M ($1.17 E-4$), G ($1.18 E-4$) e L ($1.34E-4$) são similares, com sobreposição dos intervalos de confiança.

Quando analisamos o ciclo aéreo, observamos que as médias das taxas de substituição encontradas entre os genes P ($2.28 E-4$) e M ($2.49 E-4$) são similares, porém diferentes das taxas encontradas para os genes N ($1.30E-4$), G ($1.4E-4$) e L ($1.15E-4$), as quais também são similares.

Depois de discutidos os aspectos da heterotaquia entre os genes do RABV dentro de um mesmo ciclo, discutiremos a heterotaquia encontrada relacionando-se os 5 genes dos RABV circulantes nos ciclos aéreo e terrestre, ou seja, heterotaquia interciclos.

De acordo com os resultados expostos no item 4.8, podemos observar as diferenças para as taxas de substituição de nucleotídeos por sítio por ano encontradas nos genes N, P, M, G e L do ciclo aéreo em relação a seus homólogos no ciclo terrestre do RABV (Quadro 8), sendo as maiores heterotaquias observadas para os genes N, P e M.

Nos resultados apresentados podemos observar que as diferentes taxas de substituição (heterotaquia) foram bastante evidentes para os genes N, P e M dos RABV distribuídos nos ciclos aéreo e terrestre da raiva. Embora o gene L dos RABV dos dois ciclos também tenha apresentado uma pequena diferença em suas taxa de substituição.

De acordo com os resultados obtidos, o gene N parece estar acumulando substituições de nucleotídeos a uma taxa aproximadamente duas vezes maior no ciclo terrestre ($2.26E-4$) em relação ao aéreo ($1.30E-4$), sem sobreposições de intervalos de confiança. Este resultado indica que as restrições para as substituições de nucleotídeos por sítios no gene N dos RABV circulantes no ciclo aéreo são maiores que as restrições impostas para a substituição de nucleotídeos por sítios nos RABV circulantes no ciclo terrestre.

No ciclo aéreo foram encontradas as maiores taxa de substituição para os genes P ($2.04E-4$ - $2.50E-4$) e M ($2.37E-4$ - $2.60E-4$) em relação os genes P ($6.9E-5$ - $1.64E-4$) e M ($6.79E-5$ - $1.62E-4$) dos RABV do ciclo terrestre, sem sobreposição dos intervalos de confiança. O gene L apresentou uma pequena heterotaquia, sendo que as maiores taxas foram encontradas no ciclo terrestre ($1.32E-4$ - $1.36E-4$) em relação ao ciclo aéreo ($1.13E-4$ - $1.16E-4$).

O sistema imune pode estar relacionado às diferentes taxas de substituição de nucleotídeos encontradas para os genes sob heterotquia, uma vez que as respostas imunes ao RABV são diferentes entre os quirópteros e os carnívoros. Embora se conheçam muitos aspectos das interações das proteínas N, P e G na resposta imune contra o RABV, muitos aspectos da interação das proteínas M e L com o sistema imune ainda não estão bem estabelecidos (LAFON, 2007; WUNNER, 2007; SCHNELL et al., 2010).

De acordo com as taxas de substituição de nucleotídeos encontradas para os 5 genes dos RABV nos ciclos aéreo e terrestre, podemos concluir que apenas os genes G e L estão acumulando substituições de nucleotídeos por sítio por ano em taxas aproximadas nos dois ciclos. Isto significa que análises filogenéticas que utilizem as taxas de substituições de nucleotídeos por sítio por ano para estabelecer o tempo de ancestral comum mais recente

(TACMR) não deveriam ser realizadas com o genoma completo do RABV, mas apenas com os genes que parecem estar evoluindo em velocidade similares entre as diversas linhagens em estudo.

Isto pode ser evidenciado nas diferenças entre os resultados obtidos para o TACMR para os genes N, M e P nos ciclos aéreo e terrestre do RABV. Já para os genes G e L, estes valores médios de TACMR foram próximos tanto para cada gene em ambos os ciclo, quanto para os dois genes em um mesmo ciclo (Quadro 8).

Os resultados que demonstram as grandes diferenças entre as taxas de substituições de nucleotídeos por sítio por ano existentes entre os genes N, P e M dos RABV mantidos nos ciclos aéreo e terrestre, aliados aos resultados encontrados para a análise do sinal filogenético de cada gene (Quadro 9), serviram de parâmetro para a escolha dos genes G e L concatenados para o cálculo da estimativa do TACMR entre os ciclos aéreo e terrestre do RABV.

De acordo com os resultados encontrados para o TACMR para os ciclos aéreo e terrestre do RABV, esta divergência teria ocorrido entre 1178 a 2723 anos atrás, com 95% de intervalo de confiança, sendo a idade média desta divergência em torno de 1870 anos. Na árvore filogenética de máxima credibilidade de clados (MCC), o nó de divergência entre os dois ciclos tem o valor de 1783 anos (1101 a 2937), sendo estas diferenças inerentes as inferências bayesianas (DRUMMOND et al., 2005, 2012) .

Bourhy et al. (2008) utilizando 151 sequências do gene N completo (25 sequências do ciclo aéreo e 126 sequências do ciclo terrestre) inferiu que o TACMR para os ciclos aéreo e terrestre seria de 749 anos (363-1215 anos). Neste mesmo trabalho utilizando-se de 74 sequências do gene G completo (2 do ciclo aéreo e 72 do ciclo terrestre) encontrou-se um TACMR entre os ciclos aéreo e terrestre de 583 anos (222 a 1116 anos).

Como visto até o momento, o gene N, apesar de ser o mais conservado dos genes do RABV, não fornece o melhor sinal filogenético, e também não apresenta taxas similares de substituição de nucleotídeos por sítio por ano nos ciclos aéreo e terrestre, e por tanto não é o gene ideal para este tipo de análise. Estes fatos, aliados a pouca representatividade das linhagens representativas do ciclo aéreo utilizadas por Bourhy et al. (2008) foram os prováveis responsáveis pela divergências nos resultados encontrados relativos ao TACMR dos ciclos aéreo e terrestre. Já as diferenças encontradas para os resultados apresentados por Bourhy et al. (2008) com relação ao TACMR entre os ciclos aéreo e terrestre a partir da análise do gene G e os resultados apresentados nesta tese de doutorados devem-se provavelmente a falta de representatividade amostral utilizada pelo referido autor.

A grande quantidade de sinal filogenético fornecida pela região G-L aliada grande representatividade de linhagens do RABV dos ciclos terrestre e aéreo utilizadas nesta tese de doutorado aumentam a confiabilidade dos resultados obtidos quanto ao TACMR entre estes ciclos. Isto pode ser comprovado no valor máximo de posterior neste nó (1) na árvore filogenética de máxima credibilidade de clados obtida na análise do TACMR.

Com base na árvore de MCC, o valor médio encontrado para o TACMR para cada ciclo (ancestral comum de cada ciclo) com base nos genes G-L concatenados foi de 1166 anos (712 a 1881 anos) para o ciclo aéreo (posterior de 0,99) e 1409 anos (888 a 2218 anos) para o ciclo terrestre (posterior de 0,93), indicando pelo valor médio que os representantes do RABV do ciclo terrestre seriam mais antigos e estão evoluindo há mais tempo que os RABV do ciclo aéreo, embora observamos sobreposição dos intervalos de confiança, o que é esperado, uma vez que os dois ciclos divergiram em um mesmo momento (1783 anos). A hipótese mais provável para esta diferenças médias seja a representatividade das sequências de RABV utilizadas na análise, uma vez que provavelmente os representantes mais antigos de cada ciclo não foram contemplados na mesma.

Ainda analisando os agrupamentos basais de cada ciclo, observa-se que todas as linhagens representativas do ciclo terrestre compartilharam um ancestral comum com a linhagem mais basal Ásia 3/Índia, linhagem que tem como reservatório o cão doméstico. O tempo médio estimado de divergência desta linhagem da demais do ciclo terrestre no presente estudo foi de aproximadamente 1409 anos (variando entre 888 e 2218 anos).

Similarmente, em análises envolvendo o gene N e G separadamente, as mesmas inferências relativas à linhagem Ásia 3/Índia como a primeira a divergir dentro do ciclo terrestre também foram apresentadas por Bourhy et al. (2008) e Kuzmin et al. (2012), porém nestes trabalhos não foram apresentadas as estimativas do tempo de divergência para esta linhagem.

No ciclo terrestre, com base no sequenciamento completo da glicoproteína viral são descritos oito grandes grupos de linhagens genéticas que tem como reservatórios apenas representantes da ordem *Carnivora*, principalmente canídeos (BOURHY et al., 2008; KUZMIN et al., 2012), sendo que na presente análise contemplaram-se sete destes grandes grupos de linhagens, com exceção da linhagem África 3. Entre estas oito linhagens identificadas para o ciclo terrestre, apenas a linhagem África 3, que tem como reservatório o mangusto amarelo africano *Cynictis penicillata*, não é mantida por canídeos. As linhagens artic e artic-like são mantidas pela raposa do ártico e raccondogs, enquanto diferentes

canídeos silvestres mantém alguns ciclos dentro do complexo de linhagens do ciclo terrestre conhecido como cosmopolita, nos Estados Unidos e na Europa.

A estimativa de origem do complexo de linhagens Cosmopolita, onde se encontram as linhagens mantidas por canídeos no Brasil, se dá por volta do ano 1300, tendo como grupo basal uma amostra isolada nos anos cinquenta em um cão de Israel (RV2324 - KF154998.1.) Em seu estudo sobre a filogeografia da raiva canina mundial, Bourhy et al. (2008) encontrou como grupo basal para o complexo cosmopolita uma amostra isolada em um ser humano nos anos 70 no Egito (8692EGY - U22627), a qual apresenta grande similaridade genética com o isolado RV2324.

Esta linhagem possui um ancestral comum com o restante das linhagens existentes no complexo cosmopolita, que compreendem uma das linhagens mantidas por canídeos silvestres na América do Sul, as linhagens mantidas por cangambás e canídeos silvestres na América do norte e algumas das linhagens mantidas por canídeos silvestres na Eurásia (EUR) e cães na África (África 1), que provavelmente representa um evento de reentrada da raiva na África a partir das Américas por volta de 500 anos.

Assim como na árvore filogenética apresentada por Kuzmin et al. (2012), os resultados encontrados nesta pesquisa demonstram que o agrupamento mais basal das linhagens conhecida como EUR e Africa 1 seria a linhagem mantida por cangambás na regiões central e norte dos EUA. Isto significa que estas duas linhagens do velho mundo teriam um ancestral comum com uma das linhagens de cangambás do ciclo terrestre do RABV nos EUA.

De acordo com a topologia da árvore filogenética baseada no gene G completo demonstrada por Kuzmin et al. (2012) o isolado brasileiro mantido pelo canídeo silvestre *C. thous* (AB362483) seria o grupo mais basal de diversas linhagens mais internas do complexo cosmopolita, o que não foi corroborado pelo resultados obtidos nesta pesquisa de tese de doutorado.

As linhagens mantidas no Brasil em ciclos epidemiológicos independentes pelo cachorro do mato (*C. thous*) e pelo cão doméstico aparentemente se originaram de um ancestral comum e estão evoluindo em cada um destes reservatórios há aproximadamente 440 anos.

No ciclo aéreo do RABV, com base no sequenciamento completo da glicoproteína viral, Kuzmin et al. (2012) descreveram a existência de 21 linhagens espécie específicas, embora Oliveira et al. (2010), considere a linhagem descrita por Kuzmin et al. (2012) como *T. brasiliensis*, como sendo duas linhagens distintas, denominadas *T. brasiliensis* South America e *T. brasiliensis* North America.

Streicker et al. (2010) baseados no sequenciamento de 372 amostras de RABV isoladas de 23 espécies de morcegos insetívoros dos EUA, descrevem a existência de 18 linhagens de RABV espécies específicas, todas elas descritas por Kuzmin et al. (2012). Além disso, outras pesquisas, baseadas no gene N, demonstram a existência de pelo menos outras 12 linhagens gênero específicas existentes em morcegos da América do Sul (KOBAYASHI et al., 2005; CUNHA, 2006; KOBAYASHI et al., 2007; OLIVEIRA et al., 2010; ALBAS et al., 2011; ALMEIDA et al., 2011; MENOZZI, 2012; QUEIROZ et al., 2012; PIÑERO et al., 2012; ALLENDORF et al., 2012; YUNG et al., 2012; FAVORETTO et al., 2013) não contempladas nos dois trabalhos de Streicker et al. (2010) e Kuzmin et al. (2012). Sendo assim, até o momento, foram descritas 34 linhagens pertencentes ao ciclo aéreo do RABV, sendo 31 delas mantidas por morcegos, sendo que alguns gêneros de morcegos, como *Myotis*, *Eptesicus* e *Lasiurus* são reservatórios de mais de uma linhagem de RABV. Os outros reservatórios de RABV do ciclo aéreo são o guaxinim e o cangambá no México e América do Norte, e o sagui do tufo branco (*C. jacchus*) no nordeste brasileiro.

Nesta pesquisa foram contempladas 26 das 34 linhagens já descritas no ciclo aéreo do RABV, sendo que as linhagens *Nyctinomops* (NY-BR), *Eptesicus* 1 (EP-BR1), *Eptesicus* 2 (EP-BR2), *Myotis* Brazil (MY-BR) e *C. jacchus* (CJ-BR) tiveram pela primeira vez todos os seus genes sequenciados (Figura 7).

Assim como já demonstrado por outros autores, na árvore filogenética de MCC inferida a partir dos genes G-L concatenados pode-se observar que as linhagens mantidas pelo cangambá mexicano (*Spilogale putorius*), pelo cangambá americano (*M. mephitis*) e pelo guaxinim norte americano (*P. lotor*) compartilham um ancestral comum com todas as outras linhagens do ciclo aéreo mantidas por morcegos e pelo sagui do tufo branco *C. jacchus* (KUZMIN et al., 2012), sendo que a divergência entre estes dois grupos se deu aproximadamente a 1160 anos. Os representantes atuais deste complexo de linhagens do ciclo aéreo mantido por carnívoros silvestres norte americanos originaram-se provavelmente há aproximadamente 840 anos. Isto significa que o vírus da raiva mantido nas Américas no ciclo aéreo teve origem por volta de 1200 anos, sendo a sua procedência ainda é um mistério.

Dentro do grande grupo de linhagens do ciclo aéreo mantidas por morcegos e pelo sagui *C. jacchus*, o grupo mais basal é composto por duas linhagens brasileiras, *Myotis* Brazil e *Eptesicus* Brazil 1, as quais compartilharam um ancestral comum por volta de 730 anos, sendo que o RABV ancestral destas duas linhagens, também é compartilhado por todas as outras linhagens mantidas por morcegos em toda a América, sendo que esta divergência está datada em aproximadamente 900 anos.

A existência de uma linhagem de RABV específica mantida pelo sagui do tufo branco *C. jacchus* no nordeste brasileiro foi descrita em 2002, em um estudo que reconheceu este animal como sendo a fonte de infecção de oito casos de raiva humana nos estado do Ceará ocorridos entre 1991 e 1998, sendo que este animal já foi responsável por 9 mortes humanas entre 2000 e 2013² naquela região (SVS/MINISTÉRIO DA SAÚDE). No entanto, a região do gene N analisada e o conjunto de dados utilizados nestes trabalhos não foram os ideais para a realização de análises filogenéticas para a inferência da origem e ancestralidade relacionadas a esta linhagem (FAVORETTO et al., 2001).

Com relação à linhagem *C. jacchus*, sua origem foi estimada há provavelmente aproximadamente 380 anos, sendo que seu ancestral comum parece ser o mesmo compartilhado com o complexo de linhagens mantido por diferentes espécies do gênero *Lasiurus* nas Américas, com as linhagens *Lasionycteris noctivagans* e *Perimyotis subflavus*, sendo que esta diversificação teria ocorrido por volta de 630 anos. Este complexo de linhagens provavelmente compartilhou um ancestral comum com a linhagens norte americana *Lasiurus intermedius*, sendo sua divergência estimada em aproximadamente 700 anos.

As linhagem *Nyctinomops* Brazil e *Eptesicus* Brazil 2 já foram descritas por outros autores (KOBAYASHI et al., 2005; KOBAYASHI, Y et al., 2007; OLIVEIRA et al., 2010; ALBAS et al., 2011; ALMEIDA et al., 2011; ALLENDORF et al., 2012), porém sua provável divergência há aproximadamente 460 anos ainda não havia sido sugerida.

A relação direta de parentesco entre as linhagens *T. brasiliensis* North America e *D. rotundus* já é conhecida, sendo que sua divergência, segundo nossos resultados, ocorreu há aproximadamente 400 anos. Como visto em outros estudos, a linhagem *T. brasiliensis* South America é provavelmente o grupo basal em relação às linhagens *T. brasiliensis* North America e *D. rotundus* (KOBAYASHI et al., 2005; OLIVEIRA et al., 2010; KUZMIN et al., 2012), sendo que sua divergência do ancestral comum que originou estas duas linhagens ocorreu provavelmente por volta de 600 anos.

Enquanto no ciclo terrestre do RABV podemos observar claramente a distribuição geográfica das amostras do RABV, seu deslocamento da Índia para a Eurásia e África, para finalmente chegar as Américas através das grandes navegações por volta do século XV, o isolamento geográfico não é tão evidenciada no ciclo aéreo, sendo que a compartimentalização das linhagens neste ciclo deve-se principalmente a especificidade de reservatórios, embora certas populações dentro de algumas linhagens do ciclo aéreo

² Dados cedidos pela Secretaria de Vigilância em Saúde do Ministério da Saúde – SVS/MS

apresentem distribuição regional. Um bom exemplo é a linhagem *L. cinereus* que já foi isolada em morcegos do gênero *Lasirus* na América do Norte e no Brasil, demonstrando a circulação intercontinental desta linhagem (STREICKER et al., 2010; OLIVEIRA et al., 2010; KUZMIN et al., 2012).

Com base na representatividade dos dados utilizados nesta pesquisa, podemos considerar a existência de 5 gêneros de hospedeiros atuando nos processos seletivos envolvidos na adaptação e seleção das linhagens representativas do ciclo terrestre (*Canis*, *Vulpes*, *Nyctereutes*, *Mephitis* e *Cynictis*) contra pelo menos 13 gêneros atuando nos processos seletivos envolvidos na seleção das linhagens representativas do ciclo aéreo do RABV (*Mephitis*, *Procyon*, *Callithrix*, *Myotis*, *Eptesicus*, *Lasiurus*, *Nyctinomops*, *Tadarida*, *Perimyotis*, *Lasionycteris*, *Desmodus*, *Parastrellus* e *Antrozous*), sendo que no ciclo aéreo, ainda temos determinadas linhagens específicas para espécies dentro de um determinado gênero (FAVORETTO et al., 2001; STREICKER et al., 2010; OLIVEIRA et al., 2010; KUZMIN et al., 2012)

A utilização de sequências genômicas nesta pesquisa representou um grande gargalo no que diz respeito à representatividade de linhagens do RABV mundialmente distribuídas, uma vez que muitos representantes destas linhagens não têm o seu genoma sequenciado. Para resultados mais confiáveis, esta análise deveria ser refeita apenas com o gene G, uma vez que este gene que fornece o melhor sinal filogenético, não está sob heterotaquia nos RABV dos ciclos aéreo e terrestre, possibilitando a utilização de mais representantes de linhagens do RABV, sendo o segundo gene mais sequenciado e estudado entre os 5 genes do RABV.

Desde o evento de divergência adaptativa de um RABV ancestral em diferentes reservatórios dando origem aos ciclos aéreo e terrestre da raiva existentes atualmente, podemos observar a adaptação do RABV em diferentes reservatórios e sua evolução nas diferentes linhagens descritas na literatura, à maioria das quais está contemplada nesta tese de doutorado (STREICKER et al., 2010; OLIVEIRA et al., 2010; KUZMIN et al., 2012)

Cada reservatório é um hábitat específico e isolado, no qual podemos observar o surgimento diferentes fenótipos de RABV resultantes da interação genótipo meio ambiente. Podemos dizer que cada reservatório é responsável por um processo análogo ao isolamento geográfico estudado na evolução clássica, necessário aos eventos de evolução e especiação das diferentes populações. Reservatórios simpátricos de espécies diferentes podem ser considerados meio ambientes isolados e diferentes entre si do ponto de vista viral, mesmo ocupado o mesmo hábitat (WASIK; TURNER, 2013)

As diferentes funções e interações específicas das proteínas do RABV no sistema biológico formado entre ele e a célula infectada estão relacionadas aos diferentes processos seletivos por trás da adaptação e evolução de diferentes linhagens do RABV em diferentes reservatórios, principalmente no ciclo aéreo.

Se considerarmos a célula infectada pelo RABV como um sistema biológico complexo e adaptativo, no qual a propriedade emergente deste sistema (o RABV), só pode ser explicada pela interação de todos os seus componentes, fica claro que a adaptação e evolução de todos os elementos deste sistema estão interligadas. Sendo assim, as adaptações e evoluções de todos os genes dos RABV estão relacionadas entre si. Além disso, a evolução das diferentes proteínas celulares relacionadas ao ciclo de vida do RABV também influenciam na adaptação e evolução dos seus genes e respectivas proteínas nos diferentes reservatórios nos quais o RABV é encontrado. Cada espécie de reservatório do RABV pode ser considerado como um habitat específico no qual cada linhagem viral se adapta e evolui, sendo que a diversidade de reservatórios encontrada atualmente para o RABV podem ser consideradas como diferentes paisagens adaptativas nas quais diferentes populações virais podem se adaptar de diferentes maneiras (WRIGHT, 1932).

5.6 ANÁLISE DO SINAL FILOGENÉTICO

Como demonstrado no item 4.9, com base no conjunto de dados utilizados, o gene que apresentou o mais confiável sinal filogenético, foi o gene G, com 99,6% de quartetos resolvidos, seguidos genoma completo (99,3%) e do gene L (98,9%) (Quadro 9).

Este resultado, juntamente com o achado da existência de heterotaquia para os genes N, P e M, demonstram que os mais confiáveis genes para a realização de análises filogenéticas que envolvam os ciclos aéreo e terrestre do RABV são os genes G e L concatenados.

Embora o genoma completo tenha se mostrado melhor que o gene da RNA polimerase viral na pesquisa do sinal filogenético (Quadro 9), a existência da heterotaquia para os genes N, P e M tornam a escolha do genoma completo do RABV inadequado para certos tipos de

análises que envolvam os ciclos aéreos e terrestres do RABV, principalmente aquelas relacionadas a estimativa de taxa de substituição e TACMR.

Talvez devido ao fato de que a maioria dos estudos antigênicos feitos na época do início das pesquisas genéticas relativas ao RABV utilizarem o gene codificador da nucleoproteína viral para suas inferências e também por ser o gene N o mais abundantemente transcrito para mRNA no RABV, este gene ou regiões dele passaram ser utilizados para a realização de estudos filogenéticos do RABV, dada a maior sensibilidade de reações de amplificação. Com a descoberta de que o gene N se apresenta mais conservado em relação ao gene G do RABV, criou-se a hipótese de que o gene N, na impossibilidade da utilização de genomas completos, seria o gene ideal para a realização de estudos filogenético com o RABV, o que só contribuiu para a consolidação deste modismo.

A amplificação e sequenciamento do gene L ainda é pouco exequível para muitos dos laboratórios de biologia molecular que trabalham com o RABV, porém, a amplificação por RT-PCR ou clonagem do gene G para o seu sequenciamento já está padronizado e descrito na literatura há muitos anos (TORDO, et al., 1986 a,b). Sendo assim, na impossibilidade da utilização dos genes G-L concatenados, pode-se dizer diante do exposto de que o gene G seria o ideal para a realização de estudos filogenético do RABV e que a preferência da utilização do gene N para estes estudos não se justifica, uma vez que os requisitos técnicos e laboratoriais para a amplificação e/ou clonagem dos genes N e G e seu posterior sequenciamento são os mesmos.

5.7 ANÁLISE DO REGIME DE SELEÇÃO

A robustez genética pode ser definida como a manutenção fenotípica na presença de mudanças genéticas. Se uma mutação é responsável pela mudança de um fenótipo, podemos considerar este gene ou genoma não robusto mutacionalmente. Um genoma é considerado robusto quando a maioria das variações de nucleotídeos não acarreta mudanças nos aminoácidos codificados por estas regiões. A robustez genética interfere na evolutibilidade de certos organismos agindo em sua habilidade de se adaptarem a variações ambientais e mudança de habitat. Isto ocorre devido ao acúmulo em determinadas posições do genoma de

mutações sinônimas ou neutras para o habitat em questão. Quando ocorre a mudança deste habitat, esta arquitetura genômica promoverá ao organismo acesso a um grande número de mutações sinônimas benéficas para a sua adaptação (DUFFY et al., 2008; DRAGHI et al., 2010; LAURING et al., 2012; WASIK; TURNER, 2013).

Além do processo da adaptação fenotípica, a adaptação e evolução das diferentes linhagens de RABV em diferentes reservatórios também estar relacionada aos vieses relacionados ao uso preferencial de alguns códons sinônimos (SHARP; LI, 1987; WRIGHT, 1990; JENKINS; HOLMES, 2003; DUFFY et al., 2008; BRANDÃO, 2013).

A maioria dos genes e organismos não utilizam os códons sinônimos uniformemente sendo que ao invés disso, certos códons sinônimos são utilizados preferencialmente, num fenômeno conhecido como viés do uso de códons. Vieses observados no uso de códons entre espécies são o resultados de diferentes forças evolutivas atuando na escolha destes códons. O uso de códons pode apresentar grande variações, não apenas entre organismos mas também dentro de determinado genoma (SHARP; LI, 1987; WRIGHT, 1990; JENKINS; HOLMES, 2003; DUFFY et al., 2008; BRANDÃO, 2013).

O uso preferencial de códons, apesar de não influenciar diretamente a sequência de aminoácidos de uma proteína, pode ter grande impacto na eficiente tradução desta proteína e processos celulares. Entretanto, os mecanismos por trás da variação no uso de códons sinônimos ainda continua pouco compreendido. Acredita-se que a preferência por certos códons sinônimos seja moldada provavelmente por um equilíbrio entre vieses mutacionais e seleção natural, mas estimar as contribuições relativas a cada um destes processos não é tão simples, principalmente porque esta relação varia entre eucariotos e procariotos (SHARP; LI, 1987; JENKINS; HOLMES, 2003; LAURING et al., 2012; BELALOV; LUKASHEV, 2013).

Os vieses mutacionais são gerados por diferentes probabilidades de diferentes tipos de substituição de nucleotídeos ocorrem ao longo de todos os genomas. No caso do uso preferencial de códons estas substituições são neutras não alterando a adaptabilidade fenotípica dos organismos (BELALOV; LUKASHEV, 2013).

Acredita-se que um dos principais fatores associados a uso preferencial de códons seja o conteúdo GC (a frequência relativa destes nucleotídeos na sequência analisada). Outras causas já descritas são a deoxicitidina metilação em dinucleotídeos CpG (C-fosfato-G) que resulta na deaminação espontânea da citosina em timina ($C \rightarrow T$). Como esta timina não é detectada pelos mecanismos de correção de mutações, quando eles existem, estes erros não são corrigidos. No genoma humano, os dinucleotídeos CG ocorrem em 21% da frequência

esperada baseada em sua frequência nucleotídica (WRIGHT, 1990; BELALOV; LUKASHEV, 2013).

Na maioria dos organismos os dinucleotídeos TA também são encontrados menos frequentemente do que o esperado baseado em sua frequência nucleotídica. Isto estaria relacionado à prevenção de sinais de parada TAA e TAG, bem como a prevenção de dinucleotídeos UA em RNAs mensageiros, os quais são susceptíveis a ação de RNAses celulares. Muitas mutações ocorrem devido ao mecanismo de reparo não randômico de metilações ocorridas após o processo de replicação do material genético (WRIGHT, 1990; LAURING et al., 2012; BELALOV; LUKASHEV, 2013).

Muitas das diferenças encontradas no uso preferencial de códons entre diferentes níveis taxonômicos são frequentemente relacionadas à vieses mutacionais. Em contraste à seleção de códons causada por vieses mutacionais, pressões de seleção podem influenciar o uso preferencial de códons sinônimos. Os vieses de códons relacionados à seleção natural podem ser específicos para genes ou ainda posições do códon no RNAm, o que estaria relacionado otimização da tradução e/ou dobradura da proteína. Estes padrões podem ser observados através da comparação de regiões codantes e não codantes em um genoma (WRIGHT, 1990; LAURING et al., 2012; BELALOV; LUKASHEV, 2013).

A eficiência traducional tem sido apontada como uma pressão seletiva para explicar aspectos da evolução do uso de códons, vieses de pares de códons e sua ordem em um gene. Em vírus de RNA restrições na estrutura de seu genoma necessárias para sua replicação e encapsidação tem sido reconhecidas também como causas do uso preferencial de códons (LAURING et al., 2012).

A pressão traducional significa a preferência por códons mais adequados para a tradução em um dado contexto e sua evidência tem sido encontrada para genes altamente expressos em bactérias. A utilização de códons raros pode estar relacionada a taxa de tradução a fim de otimizar os níveis de expressão de proteína e o dobramento correto da proteína (BELALOV; LUKASHEV, 2013).

A tradução do RNAm é controlada em diversos estágios e por vários mecanismos. A principal parte do controle ocorre no estágio de iniciação, onde os ribossomos são complexados ao RNAm, na região 5'UTR. A fase de elongação da proteína é dirigida pela estrutura secundária do RNA e o grau de adaptação da sequência codante ao pool celular de RNAs transportadores. A disponibilidade de RNAs transportadores correspondentes a diferentes códons tem sido apontada como determinantes na velocidade e acurácia da tradução de proteínas (BELALOV; LUKASHEV, 2013).

No DNA, existem padrões que são preteridos ou preferidos. Isto pode estar relacionado com elementos funcionais do DNA, como a compactação dos nucleossomos e outras variedades de distribuições de nucleotídeos ao longo do genoma (WRIGHT, 1990; LAURING et al., 2012; BELALOV; LUKASHEV, 2013).

Já na molécula de RNA tem sido proposta a seleção para a eficiência de transcrição, na qual os RNAs mensageiros compostos com os nucleotídeos mais abundantes são transcritos mais rapidamente. Neste caso os códons serão enriquecidos com os nucleotídeos mais comuns. Além disso, também existem padrões preteridos e preferidos relacionados a estrutura da molécula de RNA mensageiro (WRIGHT, 1990; LAURING et al., 2012; BELALOV; LUKASHEV, 2013).

O uso preferencial de códons também é responsável pela otimização para a minimização do tempo em que os ribossomos ficam acoplados aos RNAs mensageiros durante sua tradução. Códon evoluindo sob seleção positiva apresentam grande correspondência aos RNAs transportadores mais abundantes em seu microambiente que possivelmente se ligam aos ribossomos mais rapidamente (WRIGHT, 1990; LAURING et al., 2012; BELALOV; LUKASHEV, 2013).

Como descrito no item 4.10, todos os genes dos RABV relacionados ao ciclo aéreo mostram-se predominantemente sob seleção purificadora. Já para o ciclo terrestre, com exceção do gene G, o qual apresentou somente o sítio 212 sob seleção positiva (posição 193 da proteína G sem o PS), os genes N, P, M e L também apresentaram com significância estatística a maioria sítios sob seleção purificadora.

A posição 193 da glicoproteína madura está localizada em um epítipo conformacional formado por duas regiões separadas, a região “neurotoxin-like” (aminoácidos 189-214) e o sítio antigênico III (aminoácidos 330-340), o qual participa da ativação da atividade fusogênica da glicoproteína do RABV, para a sua liberação no citoplasma celular (TORDO, 1996; WUNNER, 2007). A existência de seleção positiva para esta posição da glicoproteína sugere que esta substituição pode estar relacionada a capacidade de infectar mais hospedeiros ou mecanismos de escape frente ao sistema imune do hospedeiro (SCHNELL et al., 2010)., sendo que estas substituições não foram específicas para nenhuma linhagem do ciclo terrestre do RABV. No entanto, o aminoácido Alanina (A), foi encontrado apenas em representantes do complexo de linhagens cosmopolita (quadro 24).

O aminoácido mais frequente codificado pelo sítio 212 do gene G nas sequências analisadas do ciclo terrestre foi a Treonina (T), sendo que com exceção das linhagens Ásia 1 e África 2 (Figuras 7 e 8), todos outros agrupamento filogenéticos contemplado nas árvores

filogenéticas apresentadas neste trabalho apresentaram este aminoácido em algum de seus representante (Quadro 24).

Quadro 24 – Substituições encontradas para o sítio 212 do gene G no ciclo terrestre e grupos filogenéticos relacionados - São Paulo - 2014

Aminoácido	Classe	Códons utilizados	Numero de amostras	Grupo filogenético
Treonina (T)	Polar não carregado	ACC/ACT	43 (41 ACC)	Asia 2, Asia 3, Artic, Artic-like e Cosmopolita
Alanina (A)	Apolar	GCC	8	Cosmopolita
Valina (V)	Apolar	GTC	5	Africa 2, Asia 1 e Asia 2
Isoleucina (I)	Apolar	ATC	3	Asia 3, Asia 1, Cosmopolita

No quadro 11 podemos observar que para os genes N, M, P e G, não há diferença estatisticamente significativa quanto ao número de sítios sob seleção purificadora nos ciclos aéreo e terrestre do RABV, embora o ciclo aéreo para todos os genes apresentou um maior número de sítios sob seleção.

No ciclo aéreo do RABV o sítio 212 do gene G apresentou um valor de $\omega < 1$, indicando apenas seleção purificadora para este aminoácido na glicoproteína, sendo que o aminoácido T foi observado em 94 isolados, enquanto o aminoácido A está presente em 5 isolados e o aminoácido I está presente em um único isolado.

Entre os isolados que apresentaram o aminoácido A para o sitio 212 do gene G, é interessante observar que os hospedeiros deste vírus da linhagem de morcego insetívoro norte americana EFW1 são quatro raposas cinzentas e um coioote amostrados em surtos de raiva mantidos por carnívoros silvestre ocorridos entre 2001 a 2009 no Arizona, indicando um provável processo de adaptação de populações destas linhagens em uma nova paisagem adaptativa (KUZMIN et al., 2012).

Quanto à análise do número de sítios sob seleção purificadora, somente para o gene L foi encontrada uma diferença estatisticamente significativa entre o número de sítios sob seleção nos ciclos aéreo e terrestre do RABV, com um maior número de sítio de seleção encontrado no ciclo aéreo. O alto grau de conservação, as fortes restrições para a substituição de aminoácidos devido à conservação da estrutura e função, aliada ao maior número de reservatórios de linhagens do RABV no ciclo aéreo em relação ao terrestre são provavelmente variáveis relacionadas à diferença no número de sítios de seleção entre ambos os ciclos do RABV (POND et al., 2005; STREICKER et al., 2010; 2012).

O fato da seleção purificadora estar atuando predominantemente em todos os genes dos RABV analisados nesta pesquisa, determinando em última análise a manutenção de determinadas posições de aminoácidos em suas proteínas, demonstra a robustez do RABV frente a possíveis variações de habitat. Isto poderia estar relacionado à grande infectividade do RABV e de sua capacidade de se adaptar em novos reservatórios e o grande número de spillover entre diferentes linhagens do RABV de ambos os ciclos (MOLLENTZE; BIEK; STREICKER, 2014).

O ciclo aéreo apresentou maior intensidade de seleção purificadora para todos os genes com significância estatística em relação ao ciclo terrestre o que pode ser responsável por uma maior robustez genética dos RABV circulantes neste ciclo.

Embora a diferença no números de sítios sob seleção purificadora entre os ciclos aéreo e terrestre para os genes N, P, M e G não serem estatisticamente significativas, o maior número de sítios sob seleção no ciclo aéreo em relação ao terrestre pode estar relacionado ao maior número de reservatórios de linhagens de RABV existentes no ciclo aéreo.

Já para o gene L, a significância estatística entre o maior número de sítios sob seleção purificadora atuando no ciclo aéreo em relação ao terrestre pode também estar relacionada à necessidade de acúmulo de um maior número de sítios sinônimos para que a replicação viral não seja comprometida num caso de variação ambiental ou mudança de hospedeiro.

O gene L representa 54% do genoma do RABV e a RNapolimerase RNAdependente traduzida por ele é uma proteína imprescindível para a transcrição e replicação viral. Talvez estas características possam ter levado à necessidade dos RABV pertencentes ao ciclo aéreo a acumularem mais substituições sinônimas em relação ao ciclo terrestre, a fim de manterem a estrutura e função desta proteína mesmo em condições ambientais adversas.

No quadro 12 e figura 7, podemos observar que os valores de mediana de ω são sempre maiores para todos os genes do ciclo aéreo em relação ao ciclo terrestre do RABV, indicando com significância estatística, que embora os dois ciclos estejam sob seleção purificadora, a intensidade desta seleção é mais intensa no ciclo aéreo para todos os genes do RABV. Isto significa que determinados sítios sob seleção no ciclo aéreo do RABV apresentam um maior número de substituições sinônimas para determinados códons em relação ao o ciclo terrestre.

Se por um lado, com exceção para o gene L, não existam diferenças significativas quanto ao número de sítios sob seleção purificadora, quando analisamos todos os genes virais quanto à intensidade desta seleção nos dois ciclos, observamos claramente a maior intensidade da seleção purificadora nos genes dos RABV do ciclo aéreo em função

provavelmente da maior diversidade de reservatórios responsáveis pela manutenção deste ciclo em relação ao ciclo terrestre.

Diante dos resultados encontrados, podemos inferir com base em nossa análise que provavelmente diversos sítios de importância para a adaptação e evolução das linhagens de RABV analisadas estão sob seleção purificante a fim de manter os mesmos aminoácidos nestas posições das proteínas do RABV.

Com base nos resultados da análise do regime de seleção atuante nos cinco genes dos RABV estudados nesta pesquisa, observamos que os RABV mantidos nos ciclos aéreo e terrestre da raiva evidenciam a robustez genotípica para todos os 5 genes de seus genomas (DRAGHI et al., 2010; LAURING et al., 2012; WASIK; TURNER, 2013), sendo que os RABV pertencentes ao ciclo aéreo são relativamente mais robustos que os seus parentes do ciclo terrestre.

Isto pode estar relacionado diretamente com a diferença da quantidade de reservatórios encontrados nos ciclos aéreo (13 gêneros) e terrestre (5 gêneros) da raiva. Se os RABV do ciclo aéreo forem realmente mais robustos geneticamente que os do ciclo terrestre, teoricamente os RABV circulantes no ciclo aéreo estariam mais adaptados em realizar a deriva de reservatórios em relação os RABV do ciclo terrestre. Esta maior robustez genotípica dos isolados do ciclo aéreo podem estar relacionados aos diversos casos de spillover de diferentes linhagens do ciclo aéreo entre reservatórios de ambos os ciclos em todo continente americano. Tal robustez parece ser tão eficiente que já foram descritos surtos mantidos por cangambás nos EUA originados pela exatamente pela linhagem EFW1 (KUZMIN et al., 2012; MOLLENTZE et al., 2014).

Não podemos esquecer que diferenças inerentes aos sistemas imunes das ordens *Chiroptera* e *Carnivora* tem grande influencia nos processos seletivos atuantes na adaptação e evolução dos RABV na natureza (SCHNELL et al., 2010; BLACKWOOD et al., 2013).

5.8 ANÁLISE DO USO PREFERENCIAL DE CÓDONS

5.8.1 Índice de Adaptação de Códon

Os resultados encontrados relativos ao índice de adaptação de códon mostram que os cinco genes dos RABV distribuídos nos ciclos aéreo e terrestre estão bem adaptados quanto sua expressão em seus reservatórios, quando comparados a um conjunto de códon sabidamente bem adaptados em *Mus musculus*.

Este resultado dever ser considerado com ressalvas, uma vez que não utilizamos genes de alta expressão específicos para os reservatórios encontrados nos ciclos aéreo e terrestre do RABV. Neste caso, o ideal seria utilizar para esta análise genes de alta expressão em quirópteros para a estimativa dos valores de CAIs para os 5 genes do RABV no ciclo aéreo, e genes de alta expressão em canídeos para a estimativa dos valores de CAI.

Como já discutido no item 4.10, a seleção purificadora está atuando com maior intensidade nos 5 genes dos RABV circulantes no ciclo aéreo em relação ao ciclo terrestre. Isto pode estar diretamente relacionado aos maiores valores de CAI observados no ciclo aéreo em relação ao ciclo terrestre, utilizando como parâmetro genes de alta expressão de *Mus musculus*. Sendo assim, podemos inferir que os genes N, P, G e L dos RABV mantidos no ciclo aéreo seriam mais expressos em *Mus musculus* em relação aos mesmos genes expressos nos RABV mantidos nos reservatórios do ciclo terrestre, devido a sua maior robustez genética.

Para o gene M o maior valor de mediana foi encontrado no ciclo aéreo, mas sem diferença estatisticamente significativa com o valor encontrado no ciclo terrestre, não permitindo desta maneira relacionar este resultado ao resultado da maior pressão de seleção purificadora atuante neste gene do RABV do ciclo aéreo.

Valores de CAI acima de 0,6 geralmente estão relacionados a genes de alta expressão em diferentes organismos (SHARP; LI, 1987), sendo assim, independentemente dos maiores valores de CAI encontrados no ciclo aéreo em relação ao terrestre, os resultados obtidos podem significar a grande adaptação dos RABV a seus respectivos reservatórios em cada ciclo, do ponto de vista de sua expressão nas células hospedeiras.

5.8.2 Número efetivo de códons utilizados (Nc) e conteúdo GC3

Quanto à variação dos valores de mediana do Nc (52,9685 - 56,876), observamos pouco viés para de utilização do número efetivo de códons sinônimos, uma vez que tais viéses seriam caracterizados por valores de Nc abaixo de 40 (BELALOV; LUKASHEV, 2013).

O menor valor encontrado para o Nc foi para o gene M do ciclo aéreo (47,718), enquanto os maiores valores de Nc (61) foram encontrados para o gene M do ciclo aéreo e genes P de ambos os ciclos

É interessante observar que os valores de Nc iguais a 61 indicam total ausência de viés para a utilização de códons sinônimos, o que significa que os gene M do ciclo aéreo e genes P de ambos os ciclos utilizam todos os códons sinônimos possíveis para a síntese de aminoácidos (59 códons) (WRIGHT, 1990; BELALOV; LUKASHEV, 2013; BRANDÃO, 2013).

Os valores de Nc encontrados mostram que todos os genes do RABV, seja no ciclo aéreo ou terrestre, utilizam muitos ou até mesmo todos os códons sinônimos para codificar um mesmo aminoácido. Apesar de todos os genes do RABV utilizarem pelo menos dois códons sinônimos para codificar um mesmo aminoácido, as análises de RSCU indicam que alguns códons sinônimos são utilizados preferencialmente no ciclo aéreo e/ou no ciclo terrestre do RABV (Quadro 16). Isto pode estar relacionado às pressões de seleções para a eficiência da tradução das proteínas e manutenção da estrutura e função do RNAm relacionada a posição de diferentes códons sinônimos em sua molécula (COLEMAN et al., 2008; LAURING et al., 2012; BELALOV; LUKASHEV, 2013)

Em organismos procariontes existe uma relação inversamente proporcional entre Nc e expressão gênica, sendo que os genes mais expressos nestes organismos apresentam os menores valores de Nc quando comparados aos valores de Nc encontrados para os genes de menor expressão. Em eucariotos esta relação entre Nc e expressão gênica não foi encontrada. (WRIGHT, 1990). Porém, deve-se lembrar de que a expressão gênica do RABV é controlada também por fatores intrínsecos do próprio RABV, como propriedades inerentes a RNA-polimerase RNA dependente e a a própria estrutura do genoma, como por exemplo, o tamanho das regiões intergênicas (FINKE et al., 2000; SCHNELL et al., 2010).

Quando se analisa do ponto de vista da teoria dos sistemas a interação entre o RABV e um tecido ou células eucariontes infectada, o que se observa é sistema complexo e adaptativo

eucarionte, e desta forma espera-se que os resultados encontrados para os valores de Nc do RABV sigam os padrões de seus hospedeiros.

Uma comparação gráfica do Nc vs GC3 fornece uma visualização da relação do uso preferencial de códons sinônimos com vieses mutacionais ou seleção natural. Na hipótese de ausência de seleção, o gene em questão estaria sujeito a restrições causadas pelo conteúdo GC3, e os pontos no gráfico estariam em cima da curva, sendo os pontos fora da curva, indicativos de Nc relacionado à seleção natural (WRIGHT, 1990; BRANDÃO, 2013).

Sendo assim podemos observar que para os genes G e L (Figuras 11E e 11F), todas as amostras analisadas, nos ciclos aéreo e terrestre tem o seu Nc fora da curva GC3, mostrando provavelmente que o número efetivo de códons utilizados nestes genes do RABV foi moldado via seleção natural.

Para o gene N, a maioria absoluta das amostras apresentou esta mesma relação entre Nc e seleção natural, com exceção aparentemente de uma amostra circulante no ciclo terrestre. Destacadamente, observou-se pouca sobreposição das distribuições de Nc vs GC3 para o gene N, sendo que as diferenças nos conteúdos GC3 encontradas para este gene nos ciclos aéreo e terrestre poderiam explicar esta diferença de padrão na distribuição Nc vs GC3 entre os ciclos (Figura 11B e Quadro 15).

No gene P observamos que apenas amostras do ciclo aéreo se localizaram sobre a curva da frequência do conteúdo GC3, indicando os prováveis vieses mutacionais correlacionados ao conteúdo GC3 ao Nc encontrado para estas amostras de RABV.

Já para o gene M, encontramos amostras do ciclo aéreo e terrestre sobrepostas a curva do conteúdo GC3, sugerindo que este seria o gene do RABV em ambos os ciclos com a maior correlação entre Nc e vieses mutacionais.

Diante do exposto pode-se sugerir que para ambos os ciclos do RABV, a seleção natural apresenta um papel fundamental para os altos valores de Nc encontrados para todos os genes, principalmente N, G e L.

Pode-se também relacionar os resultados obtidos para os altos valores de Nc encontrados para todos os genes nos ciclos aéreo e terrestre do RABV aos resultados encontrados nas análises do regime de seleção atuante em cada gene. Os valores de Nc sugerem a grande utilização de códons sinônimos o que estaria relacionado à grande quantidade de códons em ambos os ciclos do RABV que se mostraram sob seleção purificadora.

Com relação ao ciclo aéreo, para os genes P, M e L foi encontrada significância estatística para os maiores valores de mediana de NC encontrados em relação ao ciclo

terrestre, o que pode ser relacionado aos maiores valores de dN-dS para estes genes no ciclo aéreo em relação ao ciclo terrestre, indicando novamente uma provável maior robustez genética do ciclo aéreo em relação ao ciclo terrestre (DRAGHI et al., 2010; LAURING et al., 2012; WASIK; TURNER, 2013).

5.8.3 Uso relativo de códons sinônimos (RSCU)

Como observado no quadro 16, encontraram-se diferenças estatisticamente significativas ($p < 0,05$) entre a proporção dos códons preferidos utilizados pelos cinco genes do RABV de acordo com o ciclo analisado. Isto demonstra que embora os cinco genes dos RABV dos ciclos aéreo e terrestre utilizem mais de um códon sinônimo para a tradução de seus aminoácidos, sua frequência de utilização não é mesma para cada gene em ambos os ciclos, existindo códons preferidos e preteridos dependendo do ciclo do RABV analisado.

Como já discutido, restrições relativas a manutenção da estrutura da molécula de RNAm, seleção para a otimização da tradução de proteínas e a correlação com o conjunto de RNAs transportadores mais utilizados pela célula hospedeira estão em teoria atuando nos processos adaptativos que levaram a esta divergência na proporção da utilização de códons sinônimos entre os RABV dos ciclos aéreo e terrestre (JENKINS; HOLMES, 2003; TULLER et al., 2010; LAURING et al., 2012).

5.9 DIVERSIDADE GENÉTICA CALCULADA ATRAVÉS DA DISTÂNCIA PATRÍSTICA E SUA COMPARAÇÃO COM OS RESULTADOS DE DISTÂNCIA GENÉTICA E IDENTIDADE GENÉTICA

A distância patrística é um método de estimativa de variabilidade genética que utiliza árvores filogenéticas, diferentemente dos outros métodos que se baseiam somente na análise de alinhamentos de sequências (FOURMENT; GIBBS, 2006).

As árvores de Máxima verossimilhança (ML) fornecem as melhores estimativas de mudança evolutiva e divergência genética por diferentes modelos de substituição, porque no cálculo dos comprimentos de ramo (variação genética) consideram a sobreposição de substituições e as semelhanças entre uma sequência ou táxon e todos os outros representados no conjunto de dados (FOURMENT; GIBBS, 2006). Por isso foram utilizadas árvores de máxima verossimilhança em nossas análises.

Como já descrito anteriormente, devido à existência de heterotaquia entre os RABV circulantes no ciclo aéreo e terrestre da raiva, a maioria das análises desta pesquisa foi realizada separadamente para cada ciclo do RABV com o objetivo de encontrar diferenças de resultados ou padrões específicos relacionados a cada ciclo.

Embora tenhamos apresentado até o momento indícios da divergência adaptativa ocorrida na espécie RABV devido a sua diferente evolução nos ciclos aéreo e terrestre da raiva, também é necessário estudar alguns aspectos evolutivos do RABV como um todo, afinal, até o momento todos estes vírus estão classificados na espécie RABV, de modo que sua divergência genética deve ser tratada como tal.

Desta maneira podemos gerar inferências sobre as taxas evolutivas de cada gene na espécie RABV e ao mesmo tempo analisar mudanças nestes padrões de evolução relacionadas aos ciclos aéreo e terrestre do RABV. Sendo assim, iniciaremos este tópico da discussão analisando os resultados obtidos para a análise conjunta dos RABV.

De acordo com os valores de mediana das distâncias patrísticas observados no quadro 17 e figura 11, observamos a seguinte variabilidade genética para cada gene do RABV: $N < M < L < G < P$.

Embora o mesmo padrão de variabilidade genética foi encontrado com a utilização da identidade genética, o uso desta estimativa utilizando a método de distância genética obteve o seguinte resultado: $L < N < M < G > P$.

Nas análises de correlação par a par entre os padrões de variabilidade genética obtidos por diferentes métodos, observamos a correlação mais alta ($r^2=0,94$) entre os métodos de distância patrística e identidade genética (Quadro 23).

Estudo evolutivos com o RABV que utilizaram genomas completos em suas análises mostram outro padrão de conservação entre os 5 genes virais, sendo que nestes estudos, o gene L aparece sempre como o segundo gene menos variável ($N < L < M < G < P$) (WU et al., 2007; DELMAS et al., 2008), sendo que os padrões de conservação descritos nesta pesquisa de doutorado, onde o gene M se apresenta menos variável que o gene L, não foram encontrados por outros pesquisadores.

Para o estudo da divergência genética do RABV, não existe uma padronização metodológica, por tanto cada autor costuma utilizar para este fim, a metodologia que está mais familiarizado, ou as mais utilizadas pela comunidade científica. Sendo assim, a variabilidade entre os conjuntos de dados utilizados em diferentes pesquisas aliado aos diferentes métodos disponíveis para estimar a variabilidade genética podem levar a resultados conflitantes.

Embora nossas estimativas da variabilidade genética do RABV baseada na distância patrística de cada gene não represente o consenso sobre os resultados encontrados por vários pesquisadores, elas mantiveram o mesmo padrão de divergência e alta correlação com as estimativas calculadas a partir da identidade genética, método bastante utilizado para este tipo de inferência. Além disso, os valores de mediana e média encontrada para distância patrística e identidade genética para os genes M e L foram bem próximos, bem como seus valores máximos.

Agora discutiremos os resultados relativos à variabilidade genética dos cinco genes do RABV separadamente nos ciclos aéreo e terrestre.

Como podemos observar nos resultados apresentados, a divergência adaptativa do RABV nos ciclos aéreo e terrestre da raiva, além de causar a heterotaquia entre determinados genes nestes dois ciclos, também foi responsável por diferentes padrões de variabilidade genética para cada um dos cinco genes do RABV.

Primeiramente observamos que as medianas das distâncias patrísticas de cada um dos genes dos RABV do ciclo terrestre são maiores em relação a ciclo aéreo. Este resultado pode estar diretamente relacionado ao fato da intensidade da pressão de seleção purificadora ter se mostrado estatisticamente significativa maior em todos os genes do ciclo aéreo em relação ao ciclo terrestre (Figura 7), onde uma maior seleção purificadora seria responsável por uma menor variabilidade genética.

De acordo com os valores de mediana das distâncias patrísticas (Quadro 17 e Figura 10), podemos observar o seguinte padrão de variabilidade genética para cada gene do RABV nos ciclos aéreo e terrestre:

- Aéreo: $M < L < N < G < P$.
- Terrestre: $M < N < L < G < P$.

De acordo com estes resultados, podemos observar que o gene M, o segundo gene mais conservado nos cálculos de distância patrística para os ciclos aéreo e terrestre do RABV em conjunto, apresentou-se como o gene mais conservado quando os ciclos aéreo e terrestre foram analisados separadamente. Além disso, observamos que no ciclo aéreo, o gene L foi o segundo gene mais conservado enquanto que no ciclo terrestre, o segundo gene mais conservado foi o gene N.

Isto indica que a divergência adaptativa do RABV nos ciclos aéreo e terrestre pode estar relacionada à atuação de diferentes processos seletivos em cada gene dependendo do ciclo analisado, principalmente para os genes N, M e P, os quais como já visto, estão sob acentuada heterotaquia na comparação entre os dois ciclos.

Conceitualmente sabemos que quanto maior a pressão seletiva atuante em um gene, menor será sua variabilidade devido às restrições de substituição e vice-versa. Sendo assim é provável, porém controverso diante do já relatado, que o gene M seria o gene com maiores pressões seletiva no RABV. Também podemos observar diferentes intensidades de pressões de seleção atuantes para os genes L e N em cada ciclo do RABV, uma vez que o gene L é o segundo gene mais conservado no ciclo aéreo enquanto no ciclo terrestre o segundo gene menos variável é o gene N (DUFFY et al., 2008).

Da mesma maneira provavelmente nos dois ciclos do RABV, o gene P, por ser o menos conservado, devem estar sob menor pressão de seleção o que pode ser relacionados a um maior número de substituição por sítios, em relação aos outros genes em ambos os ciclos.

Analisando a variabilidade genética estimada a partir da identidade genética para cada gene do RABV nos ciclos aéreo e terrestre, observamos o seguinte padrão de variabilidade genética: ciclo aéreo $N < M < L < P < G$; ciclo terrestre $N < M < L < G < P$. Este resultado indica diferenças nos padrões de variabilidade para os genes G e P relacionados a cada ciclo do RABV.

Quando analisamos a variabilidade genética a partir do cálculo de distância genética, também observamos diferentes padrões para os 5 genes do RABV em cada ciclo: ciclo aéreo $L < N < M < G < P$; ciclo terrestre $N = M < L < G < P$.

Os coeficientes de correlação entre as diversidades genéticas estimadas para cada gene do RABV em cada ciclo pelos três métodos utilizado nesta pesquisa pode ser observados nos quadro 21 e 22.

Com relação ao ciclo terrestre, a maior correlação entre as divergências genéticas obtidas com os diferentes modelos utilizados foi encontrado entre a distância genética e a identidade genética ($r^2=0,99$), porém, a segunda maior correlação foi encontrada na comparação da distância patrística com a identidade genética ($r^2=0,98$).

O ciclo aéreo do RABV foi o que obteve os piores coeficientes de correlação entre o diferentes métodos utilizados para se inferir os padrões de variabilidade genética dos cinco genes.

Baseado nestes resultados, observamos uma boa correlação entre os três métodos utilizados para estimar a divergência genética nas amostras representativas do ciclo terrestre, com a melhor correlação ocorrendo entre a distância genética e identidade genética ($r^2=0,99$).

Já as correlações entre as variabilidades de cada gene estimadas para o ciclo aéreo não foram tão altas, sendo que o melhor coeficiente de correlação ($r^2=0,89$) foi obtido entre a distância patrística e a identidade genética.

Até o momento todos os estudo relacionados a filogenia e evolução do RABV foram realizados sem levar em consideração a existência da heterotaquia entre os ciclos aéreo e terrestre. Sendo assim, os efeitos da divergência adaptativa que deu origem aos diferentes padrões de variabilidade genética entre os entre os RABV distribuídos nestes dois ciclos ainda não haviam sido descritos.

As estimativas encontradas para a divergência genética do RABV nesta tese de doutorado não diferem muito das já citadas na literatura, com exceção da maior conservação da proteína M em relação à proteína L. A alta correlação encontrada entre a distância patrística e a identidade genética encontradas para os cinco genes do RABV, aliada a manutenção do mesmo padrão de divergência genética encontrado para os dois modelos indicam alta probabilidade da ocorrência deste padrão de divergência genética entre os 5 genes estudados nas amostras de RABV contempladas neste estudo.

No entanto, quando analisamos os ciclos aéreo e terrestre do RABV separadamente, observamos diferentes padrões de divergência específicos ocorrendo em cada um dos genes do RABV, relacionados a cada um dos ciclos. Este padrões diferem entre si para todos os modelos de estimativa genética utilizados, sendo que para cada modelo testado um padrão diferente foi obtido para a divergência genética em cada ciclo.

6 CONCLUSÕES

A padronização da técnica de RT-PCR genômica para o RABV, utilizando-se como alvo o RNA anti genômico mostrou-se eficaz para a amplificação do genoma do RABV para 7 linhagens gênero ou espécie específicas existentes no Brasil, uma vez que o isolado 3208/06, representante da linhagem *Eptesicus* Brazil II foi amplificada em duas reações de PCR.

Como observado, a hipótese de que os RABV dos ciclos aéreo e terrestre estão sob relógios moleculares distintos foi melhor aceita em relação a hipótese que ambos os ciclos do RABV estarem sob um mesmo relógio molecular. A partir deste achado, pode-se concluir a existência de heterotaquia entre os ciclos aéreo e terrestre do RABV, ocasionada provavelmente pela divergência adaptativa dos RABV circulantes em cada um destes ciclos e de maneira diferente em cada gene.

Baseado nos resultados encontrados na análise de melhor sinal filogenético com relação ao RABV, aliados aos resultados para a análise de heterotaquia entre os mesmos genes nos ciclos aéreo e terrestre, pode-se concluir que o gene G seria o gene ideal para esta finalidade.

De acordo com os resultados encontrados verificou-se que a seleção purificadora é predominante em todos os genes dos RABV em ambos os ciclos, sendo que a mesma atua mais intensamente em todos os genes dos vírus pertencentes ao ciclo aéreo em relação ao ciclo terrestre.

Nos RABV do ciclo aéreo, os genes N, M, G e L apresentaram uma quantidade maior de sítios sob seleção purificadora com relação aos RABV do ciclo terrestre, com significância estatística encontrada apenas para o gene L. O gene P apresentou mais sítios sob seleção purificadora no ciclo terrestre em relação ao ciclo aéreo, porém sem significância estatística.

Quanto ao estudo das taxas de substituição de nucleotídeos por sítio por ano realizados separadamente para cada gene do RABV em cada ciclo, pode-se concluir que divergências adaptativas relacionadas as diferentes funções e interações de cada proteína do RABV entre si e com diferentes moléculas da célula hospedeira, necessárias à realização do seu ciclo de vida e evasão do sistema imune, poderiam ter levado a fixação de diferentes taxas de substituição

de nucleotídeos por sítio por ano entre os genes N, P, M e L dos RABV em cada ciclo, com as maiores divergências encontradas entre os genes N, P e M.

Não foram encontrados resultados que sugerissem vieses na utilização de códons sinônimos em nenhum gene do RABV em nenhum dos ciclos devido aos altos valores de NC encontrados, porém a análise do RSCU mostrou que todos os aminoácidos das cinco proteínas do RABV são codificados preferencialmente por pelo menos um códon.

Todos os genes dos RABV circulantes em ambos os ciclos apresentaram altos alto índice de adaptação de códons, quando comparados a genes de alta expressão em *M. musculus*.

A análise de TACMR utilizando-se para esta estimativa os genes G-L concatenados encontrou que o tempo de divergência dos ciclos aéreo e terrestre do RABV a partir de um RABV ancestral há aproximadamente 1870 anos.

Baseado nos resultados obtidos para a análise da variabilidade genética utilizando-se a distância patrística, quando os RABV dos ciclos aéreo e terrestre foram analisados em conjunto, os padrões de variabilidade genética encontrados ($N < M < L < G < P$) foram diferentes dos padrões encontrados nas análises em separado para o ciclo aéreo ($M < L < N < G < P$) e para o ciclo terrestre ($M < N < L < G > P$), sugerindo que a divergência adaptativa gerada pelas pressões de seleção específicas em cada ciclo do RABV para cada um dos 5 genes moldaram de maneira diferentes os padrões de variabilidade genética dos mesmos em ambos os ciclos do RABV.

Pode-se concluir também a partir dos valores medianos de distância patrística obtidos que os genes pertencentes a maioria dos RABV circulantes no ciclo terrestre apresentam maior variabilidade genética em relação aos seus pares no ciclo aéreo.

Através dos resultados discutidos nesta tese de doutorado pode-se sugerir que os diferentes processos adaptativos e evolutivos originados a partir da separação dos RABV nos ciclos aéreo e terrestre originaram os diferentes padrões evolutivos encontrados nas comparações entre os dois ciclos. Estes resultados só foram possíveis uma vez que os ciclos aéreo e terrestres foram analisados separadamente para a sua posterior comparação, uma abordagem inédita até o momento para o estudo do RABV.

A separação do RABV nos ciclos aéreo e terrestre, ocorrida a cerca de 2000 anos segundo os resultados apresentados neste trabalho, moldou características e padrões específicos para os vírus representantes de cada um destes ciclos. No entanto este fato não é levado em consideração nas conceitualizações e descrições dos ciclos epidemiológicos da raiva, hoje divididos primeiramente em ciclo urbano e ciclo silvestre, levando-se em

consideração apenas aspectos dos reservatórios e meio ambiente, deixando de lado os aspectos inerentes ao agente etiológico. Ao integrarmos esta dicotomia genética e fenotípica do RABV aos ciclos epidemiológicos da raiva, talvez a divisão dos ciclos em urbano silvestre não seja a mais adequada, a qual poderia ser substituída por ciclo aéreo e terrestre, sendo que o ciclo terrestre ainda poderia ser subdividido em terrestre doméstico (linhagem Cão doméstico) e ciclo terrestre silvestre (linhagem *C. thous*).

REFERÊNCIAS

- ACHA, P. N.; SZYFRES, B. **Zoonoses y enfermedades transmisibles comunes al hombre y a los animales**. 3. ed. Washington: Organization Panamericana de la Salud, 2003. v. 2, p. 351-383.
- ACHA; P. N.; ARAMBULO, P. V. III. Rabies in the tropics, history and current status. , In: KUWERT, E.; MERIEUX, C.; KOPROWSKI, H.; BÖGEL, K. **Rabies in the tropics**. Berlin: Springer-Verlag, 1985. p. 343-359.
- AGRANOVICH, A.; MAMAN, Y.; LOUZOUN, Y. Viral proteome size and CD8+ T cell epitope density are correlated: The effect of complexity on selection. **Infection, genetics and evolution**, v. 20C, p. 71–77, Aug. 2013.
- ALBAS, A.; CAMPOS, A. C. DE A.; ARAUJO, D. B.; RODRIGUES, C. S.; SODRÉ, M. M.; DURIGON, E. L.; FAVORETTO, S. R. Molecular characterization of rabies virus isolated from non-haematophagous bats in Brazil. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 44, n. 6, p. 678–83, 2011.
- ALLENDORF, S. D.; CORTEZ, A.; HEINEMANN, M. B.; HARARY, C. M. A.; ANTUNES, J. M. A P.; PERES, M. G.; VICENTE, A. F.; SODRÉ, M. M.; ROSA, A. R. DA; MEGID, J. Rabies virus distribution in tissues and molecular characterization of strains from naturally infected non-hematophagous bats. **Virus research**, v. 165, n. 2, p. 119–25, May. 2012.
- ALMEIDA, M. F.; AGUIAR, E. A.; MARTORELLI, L. F.; SILVA, M. M. Laboratory diagnosis of rabies in chiroptera carried out in a metropolitan area of southeastern region Brazil. **Revista de saúde pública**, v. 28, n. 5, p. 341–4, Oct. 1994.
- ALMEIDA, M. F. DE; FAVORETTO, S. R.; MARTORELLI, L. F. A.; TREZZA-NETTO, J.; CAMPOS, A. C. D. A.; OZAHATA, C. H.; SODRÉ, M. M.; KATAOKA, A. P. A. G.; SACRAMENTO, D. R. V.; DURIGON, E. L. Characterization of rabies virus isolated from a colony of *Eptesicus furinalis* bats in Brazil. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 53, n. 1, p. 31–37, Feb. 2011.
- ARAI, Y. T.; KUZMIN, I. V; KAMEOKA, Y.; BOTVINKIN, A. D. New lyssavirus genotype from the Lesser Mouse-eared Bat (*Myotis blythi*), Kyrgyzstan. **Emerging infectious diseases**, v. 9, n. 3, p. 333–7, Mar. 2003.
- BADRANE, H.; TORDO, N. Host switching in Lyssavirus history from the Chiroptera to the Carnivora orders. **Journal of virology**, v. 75, n. 17, p. 8096–104, Sep. 2001.
- BAZINET, A. L.; ZWICKL, D. J.; CUMMINGS, M. P. A Gateway for Phylogenetic Analysis Powered by Grid Computing Featuring GARLI 2.0. **Systematic biology**, Apr. 2014.
- BELALOV, I. S.; LUKASHEV, A. N. Causes and implications of codon usage bias in RNA viruses. **PloS one**, v. 8, n. 2, p. e56642, Jan. 2013.

BLACKWOOD, J. C.; STREICKER, D. G.; ALTIZER, S.; ROHANI, P. Resolving the roles of immunity, pathogenesis, and immigration for rabies persistence in vampire bats. **Proceedings of the National Academy of Sciences of the United States of America**, v. 110, n. 51, p. 20837–42, Dec. 2013.

BORUCKI, M. K.; CHEN-HARRIS, H.; LAO, V.; VANIER, G.; WADFORD, D. A.; MESSENGER, S.; ALLEN, J. E. Ultra-Deep Sequencing of Intra-host Rabies Virus Populations during Cross-species Transmission. **PLoS Neglected Tropical Diseases**, v. 7, n. 11, p. e2555, Nov. 2013.

BOTVINKIN, A. D.; POLESCHUK, E. M.; KUZMIN, I. V.; BORISOVA, T. I.; GAZARYAN, S. V.; YAGER, P.; RUPPRECHT, C. E. Novel lyssaviruses isolated from bats in Russia. **Emerging infectious diseases**, v. 9, n. 12, p. 1623–5, Dec. 2003.

BOURHY, H.; REYNES, J.-M.; DUNHAM, E. J.; DACHEUX, L.; LARROUS, F.; HUONG, V. T. Q.; XU, G.; YAN, J.; MIRANDA, M. E. G.; HOLMES, E. C. The origin and phylogeography of dog rabies virus. **The Journal of general virology**, v. 89, n. 11, p. 2673–81, Nov. 2008.

BRANDÃO, P. E. The evolution of codon usage in structural and non-structural viral genes: The case of Avian coronavirus and its natural host *Gallus gallus*. **Virus research**, v. 178, n. 2, p. 264–271, Sep. 2013.

BRASS, D. A. Monoclonal antibody characterization of bat-rabies isolates. **Rabies in bats: natural history and public health implications**. Connecticut: Livia Press, 1994. 189–206 p.

CALISHER, C. H.; ELLISON, J. A. The other rabies viruses: The emergence and importance of lyssaviruses from bats and other vertebrates. **Travel medicine and infectious disease**, v. 10, n. 2, p. 69–79, Mar. 2012.

CAMPOS, A. C. D. A.; MELO, F. L.; ROMANO, C. M.; ARAUJO, D. B.; CUNHA, E. M. S.; SACRAMENTO, D. R. V.; ANDRADE ZANOTTO, P. M. DE; DURIGON, E. L.; FAVORETTO, S. R. One-step protocol for amplification of near full-length cDNA of the rabies virus genome. **Journal of virological methods**, v. 174, n. 1-2, p. 1–6, Jun. 2011.

CARINI, A. Sur une grande Épizootie de rage. **Annales de L'Institut Pasteur**, v. 25, p. 843–846, 1911.

CARNIELI, P.; BRANDÃO, P. E.; CARRIERI, M. L.; CASTILHO, J. G.; MACEDO, C. I.; MACHADO, L. M.; RANGEL, N.; CARVALHO, R. C. DE; CARVALHO, V. A. DE; MONTEBELLO, L.; WADA, M.; KOTAIT, I. Molecular epidemiology of rabies virus strains isolated from wild canids in Northeastern Brazil. **Virus research**, v. 120, n. 1-2, p. 113–20, Sep. 2006.

CARNIELI, P.; FAHL, W. D. O.; CASTILHO, J. G.; OLIVEIRA, R. D. N.; MACEDO, C. I.; DURYMANOVA, E.; JORGE, R. S. P.; MORATO, R. G.; SPÍNDOLA, R. O.; MACHADO, L. M.; UNGAR DE SÁ, J. E.; CARRIERI, M. L.; KOTAIT, I. Characterization of Rabies virus isolated from canids and identification of the main wild canid host in Northeastern Brazil. **Virus research**, v. 131, n. 1, p. 33–46, Jan. 2008.

CHARE, E. R. , GOULD, E. A., HOLMES, E. C. Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. **Journal of General Virology**, v. 84, n. 10, p. 2691–2703, Oct. 2003.

CHARIF, D.; THIOULOUSE, J.; LOBRY, J. R.; PERRIÈRE, G. Online synonymous codon usage analyses with the ade4 and seqinR packages. **Bioinformatics**, v. 21, n. 4, p. 545–7, Mar. 2005.

CHILDS, J.; REAL, L. A. Epidemiology. In: JACKSON, ALAN C; WUNNER, W. H. (Ed.). **Rabies**. 2. ed. San Diego: Academic Press, 2007. 123–199 p.

COLEMAN, J.; PAPAMICHAIL, D.; SKIENA, S.; FUTCHER, B.; WIMMER, E.; MUELLER, S. Virus attenuation by genome-scale changes in codon pair bias. **Science**, v. 320, n. 5884, p. 1784–1787, 2008.

CUNHA, E. M. S. **Caracterização genética de amostras do vírus da raiva isoladas de morcegos. Avaliação da Patogenicidade e proteção cruzada em camundongos**. 2006. 98 f. (Tese de Doutorado) Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo, 2006.

CUNHA, E. M. S.; SILVA, L. H. Q. DA; LARA, M. DO C. C. S. H.; NASSAR, A. F. C.; ALBAS, A.; SODRÉ, M. M.; PEDRO, W. A. Bat rabies in the north-northwestern regions of the state of São Paulo, Brazil: 1997-2002. **Revista de saúde pública**, v. 40, n. 6, p. 1082–6, Dec. 2006.

DELMAS, O.; HOLMES, E. C.; TALBI, C.; LARROUS, F.; DACHEUX, L.; BOUCHIER, C.; BOURHY, H. Genomic diversity and evolution of the lyssaviruses. **PloS one**, v. 3, n. 4, p. e2057, Jan. 2008.

DIAZ, A. M.; PAPO, S.; RODRIGUEZ, A.; SMITH, J. S. Antigenic analysis of rabies-virus isolates from Latin America and the Caribbean. **Zentralblatt für Veterinärmedizin. Reihe B. Journal of veterinary medicine**, v. 41, n. 3, p. 153–60, May. 1994.

DRAGHI, J.; PARSONS, T.; WAGNER, G.; PLOTKIN, J. Mutational robustness can facilitate adaptation. **Nature**, v. 463, n. 7279, p. 353–355, 2010.

DRUMMOND, A J.; RAMBAUT, A; SHAPIRO, B.; PYBUS, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. **Molecular biology and evolution**, v. 22, n. 5, p. 1185–92, May. 2005.

DRUMMOND, A. J.; HO, S. Y. W.; PHILLIPS, M. J.; RAMBAUT, A. Relaxed phylogenetics and dating with confidence. **PLoS biology**, v. 4, n. 5, p. e88, May. 2006.

DRUMMOND, A. J.; SUCHARD, M. A; XIE, D.; RAMBAUT, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. **Molecular biology and evolution**, v. 29, n. 8, p. 1969–73, Aug. 2012.

DUFFY, S.; SHACKELTON, L. A; HOLMES, E. C. Rates of evolutionary change in viruses: patterns and determinants. **Nature reviews. Genetics**, v. 9, n. 4, p. 267–76, Apr. 2008.

- EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic acids research**, v. 32, n. 5, p. 1792–7, Jan. 2004.
- EWING, B.; GREEN, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. **Genome research**, v. 8, n. 3, p. 186–94, Mar. 1998.
- FAVORETTO, S. R.; CARRIERI, M. L.; CUNHA, E. M. S.; AGUIAR, E. A C.; SILVA, L. H. Q.; SODRE, M. M.; SOUZA, M. C. A M.; KOTAIT, I. Antigenic typing of Brazilian rabies virus samples isolated from animals and humans, 1989-2000. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 44, n. 2, p. 91–5, 2002.
- FAVORETTO, S. R.; MATTOS, C. C. DE; MATTOS, C. A DE; CAMPOS, A C. A; SACRAMENTO, D. R. V; DURIGON, E. L. The emergence of wildlife species as a source of human rabies infection in Brazil. **Epidemiology and infection**, v. 141, n. 7, p. 1552–61, Jul. 2013.
- FAVORETTO, S. R.; MATTOS, C. C. DE; MORAIS, N. B.; ALVES ARAÚJO, F. A.; MATTOS, C. A. DE. Rabies in marmosets (*Callithrix jacchus*), Ceará, Brazil. **Emerging infectious diseases**, v. 7, n. 6, p. 1062–5, 2001.
- FINKE, S.; COX, J. H.; CONZELMANN, K. K. Differential transcription attenuation of rabies virus genes by intergenic regions: generation of recombinant viruses overexpressing the polymerase gene. **Journal of virology**, v. 74, n. 16, p. 7261–9, Aug. 2000.
- FOURMENT, M.; GIBBS, M. J. PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. **BMC evolutionary biology**, v. 6, p. 1, Jan. 2006.
- HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic acids symposium series**, v. 41, n. 95-98, 1999.
- HAN, G.-Z.; WOROBEY, M. Homologous recombination in negative sense RNA viruses. **Viruses**, v. 3, n. 8, p. 1358–73, Aug. 2011.
- HE, C.-Q.; MENG, S.-L.; YAN, H.-Y.; DING, N.-Z.; HE, H.-B.; YAN, J.-X.; XU, G.-L. Isolation and identification of a novel rabies virus lineage in China with natural recombinant nucleoprotein gene. **PloS one**, v. 7, n. 12, p. e49992, Jan. 2012.
- JACKSON, F. R.; TURMELLE, A. S.; FARINO, D. M.; FRANKA, R.; MCCracken, G. F.; RUPPRECHT, C. E. Experimental rabies virus infection of big brown bats (*Eptesicus fuscus*). **Journal of wildlife diseases**, v. 44, n. 3, p. 612–21, Jul. 2008.
- JENKINS, G. M.; HOLMES, E. C. The extent of codon usage bias in human RNA viruses and its evolutionary origin. **Virus research**, v. 92, n. 1, p. 1–7, Mar. 2003.
- KING, A. M. Q.; ADAMS, M. J.; CARSTENS, E. B.; LEFKOWITZ, E. J. (EDS.). Genus *Lyssavirus*. **International Committee on Taxonomy of Viruses**. San Diego: Elsevier Inc., 2012. 696–699 p.

KISSI, B.; BADRANE, H.; AUDRY, L.; LAVENU, A.; TORDO, N.; BRAHIMI, M.; BOURHY, H. Dynamics of rabies virus quasispecies during serial passages in heterologous hosts. **The Journal of general virology**, v. 80, n. 8, p. 2041–50, Aug. 1999.

KISSI, B.; TORDO, N.; BOURHY, H. Genetic polymorphism in the rabies virus nucleoprotein gene. **Virology**, v. 209, n. 2, p. 526–37, Jun. 1995.

KOBAYASHI, Y.; INOUE, N.; SATO, G.; ITOU, T.; SANTOS, H. P.; BRITO, C. J. C.; GOMES, A. A. B.; SANTOS, M. F. C.; SILVA, M. V.; MOTA, C. S.; ITO, F. H.; SAKAI, T. Phylogenetic characterization of rabies virus isolates from Carnivora in Brazil. **The Journal of veterinary medical science**, v. 69, n. 7, p. 691–6, Jul. 2007.

KOBAYASHI, Y.; SATO, G.; KATO, M.; ITOU, T.; CUNHA, E. M. S.; SILVA, M. V.; MOTA, C. S.; ITO, F. H.; SAKAI, T. Genetic diversity of bat rabies viruses in Brazil. **Archives of virology**, v. 152, n. 11, p. 1995–2004, Jan. 2007.

KOBAYASHI, Y.; SATO, G.; SHOJI, Y.; SATO, T.; ITOU, T.; CUNHA, E. M. S.; SAMARA, S. I.; CARVALHO, A. A. B.; NOCITI, D. P.; ITO, F. H.; SAKAI, T. Molecular epidemiological analysis of bat rabies viruses in Brazil. **The Journal of veterinary medical science**, v. 67, n. 7, p. 647–52, Jul. 2005.

KOTAIT, E.; CARRIERI, M. L.; CARNIELI JR, P. CASTILHO, J. G. OLIVEIRA, R. N.; MACEDO, C. I.; SCHEFFER, K. C.; ACHKAR, S. M. Reservatórios silvestres do vírus da raiva: um desafio para a saúde pública. **Boletim Epidemiológico Paulista**, v. 4, n. 40, p. 2–8, 2007.

KUZMIN, I. V.; SHI, M.; ORCIARI, L. A.; YAGER, P. A.; VELASCO-VILLA, A.; KUZMINA, N. A.; STREICKER, D. G.; BERGMAN, D. L.; RUPPRECHT, C. E. Molecular inferences suggest multiple host shifts of rabies viruses from bats to mesocarnivores in Arizona during 2001–2009. **PLoS pathogens**, v. 8, n. 6, p. e1002786, Jan. 2012.

KUZMIN, I. V.; ORCIARI, L. A.; ARAI, Y. T.; SMITH, J. S.; HANLON, C. A.; KAMEOKA, Y.; RUPPRECHT, C. E. Bat lyssaviruses (Aravan and Khujand) from Central Asia: phylogenetic relationships according to N, P and G gene sequences. **Virus Research**, v. 97, n. 2, p. 65–79, Nov. 2003.

LAFON, M. Immunology. In: JACKSON, A.; WUNNER, W. H. (Ed.). **Rabies**. 2. ed. San Diego: Academic Press, 2007. 489–504 p.

LAURING, A. S.; ACEVEDO, A.; COOPER, S. B.; ANDINO, R. Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. **Cell host & microbe**, v. 12, n. 5, p. 623–32, Nov. 2012.

LIU, W.; LIU, Y.; LIU, J.; ZHAI, J.; XIE, Y. Evidence for inter- and intra-clade recombinations in rabies virus. **Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases**, v. 11, n. 8, p. 1906–12, Dec. 2011.

LOCKHART, P.; NOVIS, P.; MILLIGAN, B. G.; RIDEN, J.; RAMBAUT, A.; LARKUM, T. Heterotachy and tree building: a case study with plastids and eubacteria. **Molecular biology and evolution**, v. 23, n. 1, p. 40–5, Jan. 2006.

MARSTON, D. A.; HORTON, D. L.; NGELEJA, C.; HAMPSON, K.; MCELHINNEY, L. M.; BANYARD, A. C.; HAYDON, D.; CLEAVELAND, S.; RUPPRECHT, C. E.; BIGAMBO, M.; FOOKS, A. R.; LEMBO, T. Ikoma lyssavirus, highly divergent novel lyssavirus in an African civet. **Emerging infectious diseases**, v. 18, n. 4, p. 664–7, Apr. 2012.

MARTIN, D. P.; LEMEY, P.; LOTT, M.; MOULTON, V.; POSADA, D.; LEFEUVRE, P. RDP3: a flexible and fast computer program for analyzing recombination. **Bioinformatics (Oxford, England)**, v. 26, n. 19, p. 2462–3, Oct. 2010.

MATTOS, C. A DE; FAVI, M.; YUNG, V.; PAVLETIC, C.; MATTOS, C. C. DE. Bat rabies in urban centers in Chile. **Journal of wildlife diseases**, v. 36, n. 2, p. 231–40, Apr. 2000.

MATTOS, C. A DE; MATTOS, C. C. DE; SMITH, J. S.; MILLER, E. T.; PAPO, S.; UTRERA, A.; OSBURN, B. I. Genetic characterization of rabies field isolates from Venezuela. **Journal of clinical microbiology**, v. 34, n. 6, p. 1553–8, Jun. 1996.

MEBATSION, T. Extensive attenuation of rabies virus by simultaneously modifying the dynein light chain binding site in the P protein and replacing Arg333 in the G protein. **Journal of virology**, v. 75, n. 23, p. 11496–502, Dec. 2001.

MENOZZI, B. D. **Caracterização antigênica e genotípica de isolados do vírus rábico, de quirópteros da cidade de Botucatu – SP e região Botucatu**. 2012. 72 f. (Dissertação de Mestrado) Faculdade de Medicina de Botucatu, Universidade Estadual Paulista “Júlio de Mesquita Filho,” 2012.

MOCHIZUKI, N.; KOBAYASHI, Y.; SATO, G.; HIRANO, S.; ITOU, T.; ITO, F. H.; SAKAI, T. Determination and molecular analysis of the complete genome sequence of two wild-type rabies viruses isolated from a haematophagous bat and a frugivorous bat in Brazil. **The Journal of veterinary medical science**, v. 73, n. 6, p. 759–66, Jun. 2011.

MOCHIZUKI, N.; KOBAYASHI, Y.; SATO, G.; ITOU, T.; GOMES, A. A. B.; ITO, F. H.; SAKAI, T. Complete genome analysis of a rabies virus isolate from Brazilian wild fox. **Archives of virology**, v. 154, n. 9, p. 1475–88, Jan. 2009.

MOLLENTZE, N.; BIEK, R.; STREICKER, D. G. The role of viral evolution in rabies host shifts and emergence. **Current opinion in virology**, v. 8C, p. 68–72, Jul. 2014.

OLIVEIRA, R. D. N.; SOUZA, S. P. DE; LOBO, R. S. V.; CASTILHO, J. G.; MACEDO, C. I.; CARNIELI, P.; FAHL, W. O.; ACHKAR, S. M.; SCHEFFER, K. C.; KOTAIT, I.; CARRIERI, M. L.; BRANDÃO, P. E. Rabies virus in insectivorous bats: implications of the diversity of the nucleoprotein and glycoprotein genes for molecular epidemiology. **Virology**, v. 405, n. 2, p. 352–60, Sep. 2010.

ORCIARI, L. A.; NIEZGODA, M.; HANLON, C. A.; SHADDOCK, J. H.; SANDERLIN, D. W.; YAGER, P. A.; RUPPRECHT, C. E. Rapid clearance of SAG-2 rabies virus from dogs after oral vaccination. **Vaccine**, v. 19, n. 31, p. 4511–8, Aug. 2001.

PAWAN, J. L. The transmission of paralytic rabies in Trinidad by the vampire bat (*Desmodus rotundus murinus* Wagner. **Caribbean medical journal**, v. 21, p. 110–36, Jan. 1959.

PIÑERO, C.; GURY DOHMEN, F.; BELTRAN, F.; MARTINEZ, L.; NOVARO, L.; RUSSO, S.; PALACIOS, G.; CISTERNA, D. M. High diversity of rabies viruses associated with insectivorous bats in Argentina: presence of several independent enzootics. **PLoS neglected tropical diseases**, v. 6, n. 5, p. e1635, Jan. 2012.

POND, S. L. K.; FROST, S. D. W.; MUSE, S. V. HyPhy: hypothesis testing using phylogenies. **Bioinformatics**, v. 21, n. 5, p. 676–9, Mar. 2005.

POSADA, D.; CRANDALL, K. A. The effect of recombination on the accuracy of phylogeny estimation. **Journal of molecular evolution**, v. 54, n. 3, p. 396–402, Mar. 2002.

QUEIROZ, L. H.; FAVORETTO, S. R.; CUNHA, E. M. S.; CAMPOS, A. C. A.; LOPES, M. C.; CARVALHO, C. DE; IAMAMOTO, K.; ARAÚJO, D. B.; VENDITTI, L. L. R.; RIBEIRO, E. S.; PEDRO, W. A.; DURIGON, E. L. Rabies in southeast Brazil: a change in the epidemiological pattern. **Archives of virology**, v. 157, n. 1, p. 93–105, Jan. 2012.

ROSA, E. S. T. DA; KOTAIT, I.; BARBOSA, T. F. S.; CARRIERI, M. L.; BRANDÃO, P. E.; PINHEIRO, A. S.; BEGOT, A. L.; WADA, M. Y.; OLIVEIRA, R. C. DE; GRISARD, E. C.; FERREIRA, M.; LIMA, R. J. D. S.; MONTEBELLO, L.; MEDEIROS, D. B. A.; SOUSA, R. C. M.; BENSABATH, G.; CARMO, E. H.; VASCONCELOS, P. F. C. Bat-transmitted human rabies outbreaks, Brazilian Amazon. **Emerging infectious diseases**, v. 12, n. 8, p. 1197–202, Aug. 2006.

RUPPRECHT, C. E.; HANLON, C. A.; HEMACHUDHA, T. Rabies re-examined. **The Lancet infectious diseases**, v. 2, n. 6, p. 327–43, Jun. 2002.

SACRAMENTO, D.; BOURHY, H.; TORDO, N. PCR technique as an alternative method for diagnosis and molecular epidemiology of rabies virus. **Molecular and cellular probes**, v. 5, n. 3, p. 229–40, Jun. 1991.

SATO, G.; ITOU, T.; SHOJI, Y.; MIURA, Y.; MIKAMI, T.; ITO, M.; KURANE, I.; SAMARA, S. I.; CARVALHO, A. A. B.; NOCITI, D. P.; ITO, F. H.; SAKAI, T. Genetic and phylogenetic analysis of glycoprotein of rabies virus isolated from several species in Brazil. **The Journal of veterinary medical science**, v. 66, n. 7, p. 747–53, Jul. 2004.

SCHIERUP, M. H.; HEIN, J. Consequences of recombination on traditional phylogenetic analysis. **Genetics**, v. 156, n. 2, p. 879–91, Oct. 2000.

SCHMIDT, H. A.; STRIMMER, K.; VINGRON, M.; HAESELER, A. VON. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. **Bioinformatics (Oxford, England)**, v. 18, n. 3, p. 502–4, Mar. 2002.

SCHNELL, M. J.; MCGETTIGAN, J. P.; WIRBLICH, C.; PAPANERI, A. The cell biology of rabies virus: using stealth to reach the brain. **Nature reviews Microbiology**, v. 8, n. 1, p. 51–61, Jan. 2010.

SHAPIRO, B.; RAMBAUT, A.; DRUMMOND, A. J. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. **Molecular biology and evolution**, v. 23, n. 1, p. 7–9, Jan. 2006.

SHARP, P. M.; LI, W. H. The codon Adaptation Index: a measure of directional synonymous codon usage bias, and its potential applications. **Nucleic acids research**, v. 15, n. 3, p. 1281–95, Feb. 1987.

SILVA, A. R. Isolamento de vírus rábico de morcego não-hematófago da espécie *Phyllostomus hastatus hastatus*. **Arquivos do Instituto Biológico**, v. 4, p. 115–120, 1964.

SODRÉ, M. M.; GAMA, A. R. DA; ALMEIDA, M. F. DE. Updated list of bat species positive for rabies in Brazil. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 52, n. 2, p. 75–81, 2010.

STEECE, R.; ALTENBACH, J. S. Prevalence of rabies specific antibodies in the Mexican free-tailed bat (*Tadarida brasiliensis mexicana*) at Lava Cave, New Mexico. **Journal of wildlife diseases**, v. 25, n. 4, p. 490–6, Oct. 1989.

STREICKER, D. G.; ALTIZER, S. M.; VELASCO-VILLA, A.; RUPPRECHT, C. E. Variable evolutionary routes to host establishment across repeated rabies virus host shifts among bats. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 48, p. 19715–20, Nov. 2012.

STREICKER, D. G.; TURMELLE, A. S.; VONHOF, M. J.; KUZMIN, I. V; MCCracken, G. F.; RUPPRECHT, C. E. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. **Science**, v. 329, n. 5992, p. 676–9, Aug. 2010.

STRIMMER, K.; HAESELER, A. VON. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 13, p. 6815–9, Jun. 1997.

TADDEI, V. A. Morcegos: algumas considerações sistemáticas e biológicas. **Boletim Técnico - Coordenadoria de Assistência Técnica Integral**, v. 172, p. 1–31, 1983.

TORDO, N. Characterization and molecular biology of rabies virus. In: MESLIN, F. X.; KAPLAN, M. M.; KOPROWSKI, H. (Ed.). **Laboratory techniques in rabies**. 4. ed. Geneva: WHO, 1996. 28–51 p.

TORDO, N.; POCH, O.; ERMINE, A; KEITH, G.; ROUGEON, F. Walking along the rabies genome: is the large G-L intergenic region a remnant gene? **Proceedings of the National Academy of Sciences of the United States of America**, v. 83, n. 11, p. 3914–8, Jun. 1986.

TORDO, N.; POCH, O.; ERMINE, A.; KEITH, G. Primary structure of leader RNA and nucleoprotein genes of the rabies genome: segmented homology with VSV. **Nucleic acids research**, v. 14, n. 6, p. 2671–83, Mar. 1986.

TULLER, T.; CARMI, A.; VESTSIGIAN, K.; NAVON, S.; DORFAN, Y.; ZABORSKE, J.; PAN, T.; DAHAN, O.; FURMAN, I.; PILPEL, Y. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. **Cell**, v. 141, n. 2, p. 344–54, Apr. 2010.

UIEDA, W.; HARMANI, N. M. S.; SILVA, M. M. S. Raiva em morcegos insetívoros (Molossidae) do Sudeste do Brasil. **Revista de saúde pública**, v. 29, n. 5, p. 293–297, 1995.

VELASCO-VILLA, A.; GÓMEZ-SIERRA, M.; HERNÁNDEZ-RODRÍGUEZ, G.; JUÁREZ-ISLAS, V.; MELÉNDEZ-FÉLIX, A.; VARGAS-PINO, F.; VELÁZQUEZ-MONROY, O.; FLISSER, A. Antigenic diversity and distribution of rabies virus in Mexico. **Journal of clinical microbiology**, v. 40, n. 3, p. 951–8, Mar. 2002.

VETRIVEL, U.; ARUNKUMAR, V.; DORAIRAJ, S. ACUA: a software tool for automated codon usage analysis. **Bioinformatics**, v. 2, n. 2, p. 62–3, Jan. 2007.

VIZOTTO, L. D.; TADDEI, V. A. A chave para determinação de quirópteros brasileiros. São José do Rio Preto, São Paulo. **Boletim de Ciências**, v. 1, p. 1–72, 1973.

WASIK, B. R.; TURNER, P. E. On the biological success of viruses. **Annual review of microbiology**, v. 67, p. 519–41, Sep. 2013.

WERTHEIM, J. O.; KOSAKOVSKY POND, S. L. Purifying selection can obscure the ancient age of viral lineages. **Molecular biology and evolution**, v. 28, n. 12, p. 3355–65, Dec. 2011.

WRIGHT, F. the roles of mutation, inbreeding, crossbreeding and selection in evolution. In: PROCEEDINGS OF THE SIXTH INTERNATIONAL CONGRESS ON GENETICS, 1932, **Anais...**, 1932. p. 355–366.

WRIGHT, F. The “effective number of codons” used in a gene. **Gene**, v. 87, n. 1, p. 23–9, Mar. 1990.

WU, X.; FRANKA, R.; VELASCO-VILLA, A.; RUPPRECHT, C. E. Are all lyssavirus genes equal for phylogenetic analyses? **Virus research**, v. 129, n. 2, p. 91–103, Nov. 2007.

WUNNER, W. H. Rabies Virus. In: JACKSON, A. C.; WUNNER, W. H. (Ed.). **Rabies**. 2. ed. San Diego: Academic Press, 2007. 23–69 p.

YUNG, V.; FAVI, M.; FERNANDES, J. Typing of the rabies virus in Chile, 2002-2008. **Epidemiology and infection**, v. 140, n. 12, p. 2157–62, Dec. 2012.