

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ELOIZA TELES CALDART**

**ANÁLISE FILOGENÉTICA: CONCEITOS BÁSICOS E
SUAS UTILIZAÇÕES COMO FERRAMENTA PARA
A EPIDEMIOLOGIA MOLECULAR DE DOENÇAS VIRAIS**

**Porto Alegre
2009/2**

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ELOIZA TELES CALDART**

**ANÁLISE FILOGENÉTICA: CONCEITOS BÁSICOS E
SUAS UTILIZAÇÕES COMO FERRAMENTA PARA
A EPIDEMIOLOGIA MOLECULAR DE DOENÇAS VIRAIS**

**Monografia apresentada à
Faculdade de Veterinária
da Universidade Federal do
Rio Grande do Sul como
requisito parcial para obtenção
de Graduação em Medicina
Veterinária**

Orientadora: Prof. Dra. Ana Paula Ravazzolo

Co-orientadora: Helena da Mata M.Sc

Porto Alegre

2009/2

C145a Caldart, Eloiza Teles

Análise filogenética: conceitos básicos e suas utilizações
como ferramenta para a epidemiologia molecular de doenças
virais / Eloiza Teles Caldart - Porto Alegre: UFRGS, 2009/2.

52f.; il. – Monografia (Graduação) – Universidade Federal
do Rio Grande do Sul, Faculdade de Veterinária, Comissão
de Estágio, Porto Alegre, BR-RS, 2009/2. Ana Paula
Ravazzolo, Orient., Helena da Matta, Co-orient.

1. Epidemiologia veterinária 2. Filogenias 3.
Epidemiologia molecular 4. Doenças virais I. Ravazzolo, Ana
Paula, Orient. II. Matta, Helena da, Co-orient. III. Título.

Catálogo na fonte
Preparada pela Biblioteca da Faculdade de
Veterinária da UFRGS

AGRADECIMENTOS

Agradeço inicialmente aos meus pais Jaqueline e Wilmar, e a minha irmã Ana Carolina que desde que decidi fazer o vestibular para Medicina Veterinária na Universidade Federal do Rio Grande do Sul, até hoje, me apoiaram não só financeiramente, mas também moralmente.

Agradeço a Deus pelas pedras no caminho que só engrandeceram meu aprendizado e pelos atalhos mostrados quando os mereci.

Agradeço às amigas estabelecidas na faculdade de Débora, Elisa e Juliana Grandi por toda a paciência quando precisei, por toda compreensão e por todos os momentos de alegria. Aos amigos feitos durante o estágio Bruna, Mariana, Haila e Amanda por terem feito meus dias longe de casa, da família e do amor mais leves e cheios de vida.

Dentro do grupo do Laboratório de Imunologia e Biologia Molecular da Universidade Federal do Rio Grande do Sul, gostaria de agradecer à professora Ana Paula Ravazzolo por mais essa etapa de orientação; à doutoranda Helena da Mata por todo o apoio e disponibilidade integral na co-orientação e à doutoranda Eliana Lopes pelos bons conselhos e disposição em ajudar e ouvir sempre.

Agradeço ao meu amor Marlon que certamente foi responsável nos últimos dois anos por recarregar minhas baterias quando pensava em desistir.

Agradeço a todas as pessoas, direta ou indiretamente, responsáveis por essa conquista.

LISTA DE FIGURAS

- Figura 1:** Exemplo de árvore filogenética enraizada_____15
- Figura 2:** Relações entre os diversos modelos evolutivos, adaptado de GARCIA (2007)_____18
- Figura 3:** Árvore filogenética da região VP4 das amostras de rotavirus isoladas de aves e mamíferos. Retirado de SCHUMANN *et al.* (2009)_____26
- Figura 4:** Reconstrução dos eventos de recombinação que levaram a emergência da recente gripe humana originada de suínos. Retirada de SMITH *et al.* (2009)_____27
- Figura 5:** Árvore filogenética do gene HA, retirada de CHEN (2006)_____28
- Figura 6:** Árvore filogenética da Família *Bunyaviridae* baseada na sequência do fragmento L. Retirada de KINSELLA *et al.* (2003)_____30
- Figura 7:** Árvore publicada em JAKAB *et al.* (2007) demonstrando a proximidade filogenética das amostras de hantavirus eslovena e húngaras_____31
- Figura 8:** Análise filogenética incluindo amostras de SRLV estudadas em PISONI *et al.* (2005)____32
- Figura 9:** Análise filogenética contendo amostras italianas estudadas por PISONI *et al.* (2005), segregando junto ao protótipo CAEV-CO (EUA) e demais amostras do *GenBank*_____33
- Figura 10:** Árvore filogenética retirada de KNOWLES *et al.* (1998) demonstrando a similaridade entre amostras italianas e amostra dos Países Baixos_____34
- Figura 11:** Árvore filogenética do grupo dos flavivirus, retirada de BATISTA *et al.* (2000)_____35
- Figura 12:** Reconstrução filogenética contendo sequências de SFV provenientes de diversos grupos de primatas não-humanos. Retirada de HUSSAIN *et al.* (2002)_____36
- Figura 13:** Árvore filogenética dos arenavirus, retirada de MONCAYO *et al.* (2000)_____37
- Figura 14:** Árvore filogenética representativa das linhagens do vírus da raiva circulante em canídeos silvestres e domésticos no Nordeste do Brasil. Retirada de JUNIOR *et al.* (2006)_____38
- Figura 15:** Os diferentes subtipos de FIV são demonstrados nessa árvore. Retirada de ANTUNES *et al.* (2008)_____40
- Figura 16:** Árvore filogenética construída com as sequências dos genes *env* e *pol* que mostra a divisão das linhagens virais em oito subgrupos com 5% de divergência entre eles. Retirada de BIEK (2006)_____41
- Figura 17:** Árvore não enraizada representando as relações entre SRLV e seus subtipos, incluindo amostras subtipo B1 de fazendas italianas onde houve mistura de rebanhos ovinos e caprinos. Retirada de PISONI *et al.* (2005)_____42
- Figura 18:** Árvores representando as relações filogenéticas entre coronavírus. Retirada de

LISTA DE TABELAS

Tabela 1: Vantagens e desvantagens dos métodos de inferência filogenética. Adaptado de HOLDER (2003) _____ 21

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

A	Adenina
BCoV	<i>Coronavirus</i> Bovino
C	Citosina
CAEV	Vírus da Artrite e Encefalite Caprina
CAEV-CO	Protótipo de CAEV isolado nos EUA (CAEV-Cork)
CCHF	Febre Hemorrágica Criméia-Congo
CCoV	<i>Coronavirus</i> Canino
ChHBV	Vírus da Hepatite B de Chimpanzés
CoV	<i>Coronavirus</i>
DNA	Ácido Desoxirribonucléico
DOBH	<i>Dobrava Hantavirus</i>
EMC	Encefalomiocardite
<i>env</i>	Gene que codifica proteínas do envelope
EUA	Estados Unidos da América
FIV	Vírus da Imunodeficiência Felina
G	Guanina
<i>gag</i>	Gene que codifica proteínas do capsídeo
GTR	<i>General Time-reversible</i>
HA	Gene que codifica proteína Hemaglutinina
HIV	Vírus da Imunodeficiência Humana
HKY	Hasegawa-Kishino-Yano
JEV	<i>Japanese Encephalitis Virus</i>
M	Gene que codifica proteína da Matriz
MEGA	Software utilizado para reconstruções filogenéticas
MP	Máxima Parcimônia
mRNA	Ácido Ribonucléico Mensageiro
MV	Máxima Verossimilhança
MVV	Vírus Maedi Visna
NA	Gene que codifica proteína Neuraminidase
NCBI	<i>National Center for Biotechnology Information</i>
NJ	<i>Neighbor-joining</i>

NP	Gene que codifica Nucleoproteína
NS	Gene que codifica proteínas não-estruturais
NSP5	Gene que codifica proteína não-estrutural 5
PA	Gene que codifica proteína da Polimerase do Vírus Influenza
PAUP*	Software utilizado para reconstruções filogenéticas
PB1	Gene que codifica proteína da Polimerase do Vírus Influenza
PB2	Gene que codifica proteína da Polimerase do Vírus Influenza
PCR	Reação de Polimerização em Cadeia
pGEM-T	Plasmídeo utilizado como vetor para clonagem
<i>pol</i>	Gene que codifica proteínas da Polimerase
RAPD	<i>Random Amplification of Polymorphic DNA</i>
RFLP	<i>Restriction Fragment Length Polymorphisms</i>
RNA	Ácido Ribonucléico
RT	Transcrição Reversa
SARS	<i>Severe Acute Respiratory Syndrome</i>
SFV	<i>Simian Foamy Virus</i>
SIV	Síndrome da Imunodeficiência Símia
SLE	<i>Saint Louis Encephalitis</i>
SRLV	Lentivírus de Pequenos Ruminantes
T	Timina
TGEV	Vírus da Gastroenterite Transmissível de Suínos
VP4	Gene que codifica proteínas de superfície
VP6	Gene que codifica proteínas de superfície
VP7	Gene que codifica proteínas de superfície
YF	<i>Yellow Fever</i>

SUMÁRIO

1	INTRODUÇÃO	10
2	EPIDEMIOLOGIA	11
2.1	Biologia Molecular em Epidemiologia	12
2.2	Epidemiologia Molecular	12
3	TERMINOLOGIA	14
4	ANÁLISE FILOGENÉTICA	16
4.1	Alinhamento Múltiplo	16
4.2	Modelo Evolutivo	16
4.2.1	Tipos de Modelos Evolutivos	17
4.3	Métodos de Reconstrução de Filogenias Moleculares	19
4.3.1	Máxima Parcimônia	19
4.3.2	Métodos de Distância - NJ	20
4.3.3	Máxima Verossimilhança	20
4.4	Árvores Filogenéticas	22
4.5	Teste de Confiança em Topologias	22
4.5.1	<i>Bootstrap</i>	22
5	RELÓGIO MOLECULAR	24
6	IMPORTÂNCIA DA ANÁLISE FILOGENÉTICA COMO FERRAMENTA EM EPIDEMIOLOGIA MOLECULAR DE DOENÇAS VIRAIS EM VETERINÁRIA	25
6.1	Estudos de Recombinação Genética	25
6.2	Caracterização da Diversidade Molecular	30
6.3	Estudos de Dinâmica Populacional	40
6.4	Outros	42
7	CONSIDERAÇÕES FINAIS	46
	REFERÊNCIAS	47

1 INTRODUÇÃO

Após o surgimento dos primeiros indícios de epidemia de HIV, no início dos anos 80, houve um aumento notável no interesse pela emergência e evolução de doenças infecciosas, principalmente com o intuito de se conhecer quais os fatores associados ao aparecimento e reaparecimento de epidemias (PAGE & HOLMES, 1998).

Há algum tempo a forma de traçar o caminho percorrido pelas doenças infecciosas era a sorologia. Hoje em dia, a constante expansão das técnicas moleculares e de suas aplicações tem mostrado alta eficiência quando usada em epidemiologia e, como consequência, essa disciplina emergente - epidemiologia molecular - cria expectativas ao redor das respostas ou questionamentos que pode gerar.

Seqüências de DNA têm se tornado ferramentas importantes para geração de árvores filogenéticas capazes de mostrar o caminho evolutivo e epidemiológico das doenças e/ou dos agentes etiológicos das mesmas.

Para HOLDER (2003), estimar árvores filogenéticas não é uma atividade exclusivamente acadêmica, em alguns casos pode sim fazer a diferença entre a vida e a morte. Por exemplo, árvores filogenéticas foram evidências cruciais para provar a transmissão de HIV de um dentista para um de seus pacientes. Além de mostrarem que casos de encefalites em Nova York representaram os primeiros exemplares de transmissão por mosquito do Vírus da Encefalite do Oeste do Nilo no Ocidente.

Estudos como os citados anteriormente serão utilizados para exemplificar de que forma a análise filogenética é uma ferramenta para epidemiologia molecular, nessa monografia.

2 EPIDEMIOLOGIA

A epidemiologia é uma ciência que estuda as condições de saúde e/ou ocorrência de doenças nas populações, procurando identificar os fatores que influenciam essas condições e essas ocorrências, tornando possível a atuação sobre eles, visando à melhoria das condições de saúde ou à prevenção das doenças. Segundo LESER (1985), a epidemiologia avalia de forma quantitativa a distribuição dos fenômenos de saúde ou doença, seus fatores condicionantes e determinantes nas populações. SOUNIS (1985) afirma que a epidemiologia permite, ainda, a avaliação da eficácia das intervenções realizadas no âmbito da saúde pública.

A análise de determinação causal das doenças em uma coletividade, dividida em classes sociais e/ou grupos específicos de populações exige da epidemiologia uma interação transdisciplinar e estabelece sua dependência de outras ciências, a exemplo das Ciências Sociais, da Ciência Política, da Estatística, da Economia, da Demografia, da Ecologia e da História (LESER, 1985).

Para FORATTINI (2005), o estudo do processo saúde ou doença, é responsável pela grande cisão da epidemiologia moderna em epidemiologia social e a epidemiologia clínica. Sendo a epidemiologia clínica a aplicação de métodos epidemiológicos à prática clínica e a epidemiologia social, a ciência que responde às demandas da medicina preventiva, promoção da saúde e necessárias intervenções sócio-econômicas para redução da pobreza, melhoria das condições de vida e saneamento do ambiente.

Segundo DORMAN (2000), a aplicação no planejamento de serviços de saúde tem sido o maior uso da epidemiologia e, em função disso, tem-se desenvolvido a legislação, as estratégias da Vigilância Epidemiológica e a organização dos sistemas de informação em saúde no âmbito governamental. O mesmo autor afirma que dentre as principais subdivisões da epidemiologia podemos destacar as avaliações de impacto ambiental, a epidemiologia genética com suas distinções dos clássicos estudos da frequência de genes mutantes e genética de populações, a epidemiologia molecular e a análise dos fenômenos biológicos ao nível celular, entre outros.

A epidemiologia em veterinária proporciona a detecção de mudanças nos fatores que interferem na distribuição de determinadas enfermidades, com a finalidade de recomendar e adotar as ações de prevenção e controle adequadas; seja para doenças importantes para a produção animal, como para a criação de animais de companhia ou para doenças com potencial zoonótico.

2.1 **Biologia Molecular em Epidemiologia**

A ocorrência periódica de mutações nos genomas leva a um fenômeno chamado polimorfismo, o qual na maior parte das vezes é silencioso; ou seja, não leva a modificações na proteína. Em alguns momentos, essas pequenas diferenças podem ser utilizadas na distinção de determinadas espécies e linhagens. Tornando-se ferramenta muito útil na sistematização e reconhecimento de organismos próximos.

A biologia molecular fornece técnicas que podem apresentar grande sensibilidade, pois são capazes de detectar e analisar a informação presente no genoma dos patógenos.

As principais técnicas de biologia molecular usadas em epidemiologia, para reconhecimento de polimorfismos, são: PCR (*polymerase chain reaction*), sequenciamento, RFLP (*restriction fragment length polymorphisms*), bancos genômicos e RAPD (*Random Amplification of Polymorphic DNA*) segundo DE ROBERTIS (2006).

2.2 **Epidemiologia Molecular**

Segundo FORATTINI (2005), a epidemiologia molecular é a ciência que foca na contribuição do potencial genético e dos fatores de risco ambientais na etiologia, na distribuição e na prevenção de doenças.

É um ramo da epidemiologia que associa informações e ferramentas da genética molecular, da clínica e da pesquisa epidemiológica clássica para definir problemas de saúde e soluções para determinadas regiões. Ou seja, estuda a evolução molecular de determinado patógeno/gene, inserido em um contexto epidemiológico (FORATTINI, 2005). Utiliza, muitas vezes, a hipótese filogenética para inferir como ocorreu a infecção na população: porque e de que forma se disseminou.

Para DORMAN (2000), o progresso dos estudos nas áreas da biologia molecular e genética está modificando a prática da medicina e da saúde pública através do desenvolvimento de diagnósticos moleculares e das intervenções direcionadas a indivíduos susceptíveis identificados geneticamente. Além disso, a epidemiologia molecular está diretamente relacionada com epidemiologia genética e essa proximidade representa a incorporação da bioquímica e biologia molecular nas pesquisas de epidemiologia tradicional, propondo o acesso a marcadores moleculares.

Para OLIVEIRA (2009), os principais objetivos dessa nova ciência são: estudos descritivos e analíticos para estimar interações hospedeiro/ambiente nas doenças; o desenvolvimento de estratégias para o controle de desordens bacterianas, parasitárias e virais através de diagnóstico molecular; e a prevenção de doenças não transmissíveis e desordens genéticas através da previsão do risco e identificando indivíduos susceptíveis através de triagem genética.

LEVIN *et al.* (1999) definem que os objetivos práticos da epidemiologia molecular são identificar os microrganismos responsáveis por doenças infecciosas, determinar sua rota de transmissão e identificar os genes responsáveis por sua virulência, resistência a drogas e produção de antígenos relacionados a vacinas.

Assim sendo, a epidemiologia molecular é uma disciplina emergente que se refere essencialmente à aplicação dos princípios da biologia molecular aos estudos epidemiológicos e, ainda, uso dos instrumentos e perspectivas epidemiológicos para compreensão das análises moleculares.

3 TERMINOLOGIA

A história dos organismos é formada por uma série de mudanças evolutivas que têm a peculiaridade de serem **hereditárias** e, portanto, derivadas da ancestralidade.

Diferentes **mecanismos de variabilidade genética** são os responsáveis pela biodiversidade existente. Esses mecanismos, segundo VANDAMME (2003), incluem mutações, duplicação de genes, reorganização de genomas e recombinação. De todos esses, principalmente as mutações, deleções e inserções são usadas nos diferentes métodos de filogenia para inferir quais as relações entre os genes.

Mutações podem ser a simples troca de uma base nitrogenada por outra, sendo **transições** mutações entre duas purinas (A e G) ou duas pirimidinas (C e T), enquanto que **transversão** é a mutação de uma purina por uma pirimidina e vice-versa. As **deleções** e **inserções** consistem na retirada de bases nitrogenadas já existentes ou fixação de novas bases ao longo do gene, respectivamente.

Essas mutações, também chamadas **polimorfismos**, podem ser **silenciosas** ou **não-silenciosas**. O primeiro causa alteração de seqüência nucleotídica no códon que codifica o aminoácido formador da proteína, mas não causa alteração no aminoácido no momento da tradução; enquanto que o segundo causa alteração no códon e no aminoácido. O **códon** é uma seqüência de três bases nitrogenadas no mRNA, os quais especificam determinado aminoácido ou indicam um ponto de parada da tradução.

A **recombinação** é um mecanismo muito importante de variabilidade genética, que consiste na troca de material genético que pode ocorrer durante a meiose ou entre microrganismos presentes no mesmo hospedeiro.

O termo **homologia** é usado quando o caractere herdado é de um ancestral comum. Por outro lado, diversos eventos, como substituições consecutivas, deleções e inserções, podem levar a **homoplasia**, que consiste em seqüências similares sem ancestral comum.

Uma **árvore filogenética (Figura 1)** é uma representação gráfica de relações ancestral-descendente entre organismos ou seqüências genéticas. As seqüências estão nas pontas (**típs**) das árvores, enquanto que os ramos (**branch**) conectam as seqüências aos seus ancestrais (HOLDER, 2003). Árvores podem ser enraizadas ou não; sendo que, as que possuem **raiz** contemplam uma seqüência chamada **outgroup**, cuja função é dar à árvore uma direção evolucionária, mostrando quem são as seqüências mais antigas (PAGE & HOLMES, 1998). Os **nós** de uma árvore filogenética representam um ancestral hipotético e são o ponto de origem dos ramos. Os diferentes arranjos dos ramos da árvore podem dar origem a diferentes **clusters**. Um **cluster**, numa definição genérica, é um

grupo de coisas ou de atividades semelhantes que se desenvolvem conjuntamente. No caso de árvores filogenéticas são sequências com certo grau de similaridade de nucleotídeos que se agrupam.

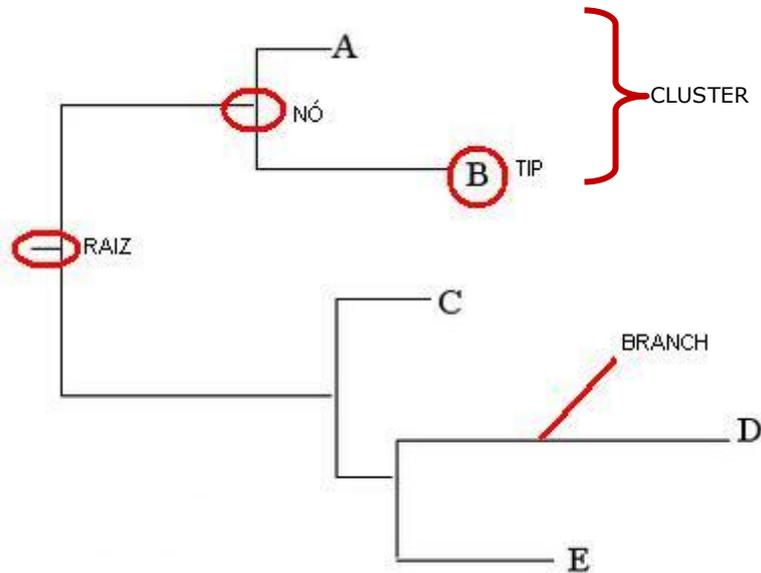


Figura 1: Exemplo de árvore filogenética enraizada.

A **distância evolucionária** entre duas sequências é o número de mudanças que ocorreram ao longo do ramo da árvore que as representam. Diferentes genes podem ter diferentes **taxas evolutivas** que, por definição, é o número de substituições de nucleotídeos por sítio, por ano ou milhões de anos.

As diversas aplicações das filogenias são taxonomia, análise de diversidade biológica, análise evolutiva de microrganismos e seus hospedeiros. A **taxonomia** é a parte da biologia voltada para a caracterização da diversidade da vida e organização do conhecimento a respeito dessa diversidade (através da estimação de relações filogenéticas entre esses organismos) (HOLDER, 2003). **Táxon** é uma unidade taxonômica essencialmente associada a um sistema de classificação; no entanto, pode estar em qualquer nível da classificação. O que significa dizer que podem ser um reino, família, gênero, uma espécie. Ou seja, entidades que de alguma maneira podem ser discernidas umas das outras. Um **sítio** corresponde ao local onde se encontra um determinado nucleotídeo na molécula de DNA.

Genes originados de um evento de duplicação recente são chamados **parálogos**. Genes homólogos em diferentes espécies, ou seja, que iniciaram uma evolução separada, são chamados **ortólogos**. Os **marcadores moleculares**, em estudos de epidemiologia molecular de doenças infecciosas, são genes ou sequências empregados para rastrear a transmissão de linhagens específicas de agentes etiológicos. Esses dados são normalmente aplicados para descrever a distribuição de

cepas em populações e para verificar os fatores de risco de hospedeiros e parasitas na dispersão da doença (GARCIA, 2007).

4 ANÁLISE FILOGENÉTICA

Sistemas de classificação de organismos têm sido organizados desde a Antiguidade e, desde essa época, já se sabe que é necessário seguir critérios que não variem de pesquisador a pesquisador, evitando a geração de informações conflitantes.

E como saber qual o melhor critério a escolher? Após a proposição da teoria da evolução por Darwin, estabeleceu-se como critério a ser escolhido as relações de ancestralidade entre espécies. Sistemas de classificação que levam em conta a evolução entre as espécies podem ser representados através de reconstruções filogenéticas e a qualidade dessa estimativa vai depender da qualidade da amostragem e das sequências amostradas (PAGE & HOLMES, 1998).

Além da classificação de organismos, análises filogenéticas têm, segundo HOLDER (2003), diversas aplicações e entre elas estão: (1) detecção de ortologia ou paralogia, (2) estimação de tempo de divergência, (3) reconstrução de proteínas fósseis, (4) procura por polimorfismos importantes na seleção natural, (5) detecção de pontos de recombinação, (6) identificação de mutações relacionadas a distúrbios ou doenças, (7) determinação da identidade de novos patógenos.

4.1 Alinhamento Múltiplo

Após a obtenção das sequências a serem utilizadas em uma reconstrução filogenética, o alinhamento é o primeiro passo para a organização das mesmas.

Um alinhamento múltiplo busca homologia posicional entre bases ou aminoácidos (SCHNEIDER, 2003). Ele pode ser realizado posicionando-se as sequências uma embaixo das outras e introduzindo “espaços” (*gaps*) para representar as inserções/deleções. Minimizando-se, assim, a diferença entre elas.

Nem todas as substituições ocorridas, entretanto, serão evidentes no alinhamento. Múltiplas substituições no mesmo sítio de uma sequência mascaram estados intermediários pelos quais a sequência passou, agravando a dificuldade de julgamento entre afinidade e convergência. Essas questões tornam a reconstrução da árvore verdadeira necessariamente uma atividade de inferência.

4.2 Modelo Evolutivo

Para construção de uma filogenia a partir de uma sequência de aminoácidos ou nucleotídeos é necessário construir uma matriz de distância baseada em um modelo de substituição de nucleotídeos ou aminoácidos, também chamados modelos evolutivos.

Há uma série de modelos evolutivos disponíveis a serem considerados, dos mais simples aos mais complexos. Essa progressão na complexidade é caracterizada pelo aumento do número de

parâmetros considerados nas análises. Os modelos mais simples avaliam a diferença relativa entre as seqüências comparadas par a par. Por outro lado, existem modelos que corrigem mutações sobrepostas, outros que consideram também as taxas de transição e transversão, diferenças quanto ao teor CG, heterogeneidade das taxas de substituições ao longo dos sítios, etc (PAGE & HOLMES, 1998). Os modelos de substituição de aminoácidos descrevem a probabilidade de mudanças de um aminoácido por outro e, portanto, são ferramentas indispensáveis para caracterização do processo de evolução de proteína.(ABASCAL *et al.*, 2005)

Na escolha do melhor modelo deve-se procurar um balanço entre acurácia e simplicidade (ABASCAL *et al.*, 2005). Portanto, os modelos mais simples devem ser, sempre que possível, preferidos aos mais complexos, pois o aumento no número de parâmetros implica no aumento da variância dos cálculos associados ao modelo (MATIOLI, 2001).

4.2.1 Tipos de Modelo Evolutivos

O modelo de substituição distância p é o mais simples e representa a proporção de posições em que as duas seqüências diferem. Apresenta a menor variância dentre os modelos e deve ser usado quando a taxa evolutiva entre as seqüências for constante (MATIOLI, 2001).

Já o modelo JUKES-CANTOR (1969 *apud* LI & GRAUR, 1991) assume que as substituições são múltiplas e ocorrem aleatoriamente entre os quatro tipos de nucleotídeos, e que os eventos de substituição obedecem à distribuição de probabilidades de Poisson. Este modelo assume igualdade nas probabilidades de substituição entre os diferentes nucleotídeos. No entanto, observa-se que transições são mais freqüentes que transversões. Isto pode ser atribuído à considerável diferença no tamanho entre pirimidinas (um anel aromático) e purinas (dois anéis aromáticos), que pode impor restrições de caráter espacial às transversões.

O modelo KIMURA 2-parâmetros (1980 *apud* RUSSO, 1995) permite levar em conta as substituições múltiplas e, justamente essa desigualdade de ocorrência entre transições e transversões, através do uso de uma taxa P de substituições transicionais e uma taxa Q de substituições transversionais.

O modelo TAJIMA-NEI (1984 *apud* KUMAR *et al.*, 2005) considera a freqüência de C e G nas seqüências do alinhamento. Este modelo não distingue transições e transversões.

O modelo TAMURA 3-parâmetros (1992) é uma proposta de extensão ao modelo Kimura 2-parâmetros, para contemplar a observação de que muitas seqüências contêm um viés muito forte na composição CG de suas seqüências. As terceiras posições (terceiro nucleotídeo) dos códons são particularmente sensíveis a este viés, dado que podem abrigar muitas substituições silenciosas.

O modelo HKY (HASEGAWA *et al.*, 1985 *apud* TAMURA, 1992) é um método probabilístico que contempla a mencionada taxa entre transições e transversões, bem como leva em consideração a frequência de ocorrência de cada uma das bases nas seqüências.

TAMURA-NEI (1993), por sua vez, é um modelo que leva em conta duas taxas de transição: transição entre purinas e transição entre pirimidinas, além de levar em conta a taxa de transversões e a desigualdade na frequência de bases das seqüências. Tamura e Nei implementaram este modelo após observar que a taxa de transição entre pirimidinas ($C \leftrightarrow T$) é muito diferente da taxa de transição entre purinas ($G \leftrightarrow A$) na região de controle do DNA mitocondrial de hominídeos, e que a negligência na observação dessa diferença acarretaria em uma subestimativa da distância real (GARCIA, 2007).

O modelo GTR (YANG, 1994) é o mais complexo (**Figura 2**). Admite seis diferentes parâmetros, que são as seis possibilidades de substituição entre as bases ($A \leftrightarrow G$; $A \leftrightarrow C$; $A \leftrightarrow T$; $G \leftrightarrow C$; $G \leftrightarrow T$; $C \leftrightarrow T$), além de levar em conta as desigualdades na composição de nucleotídeos entre as seqüências.

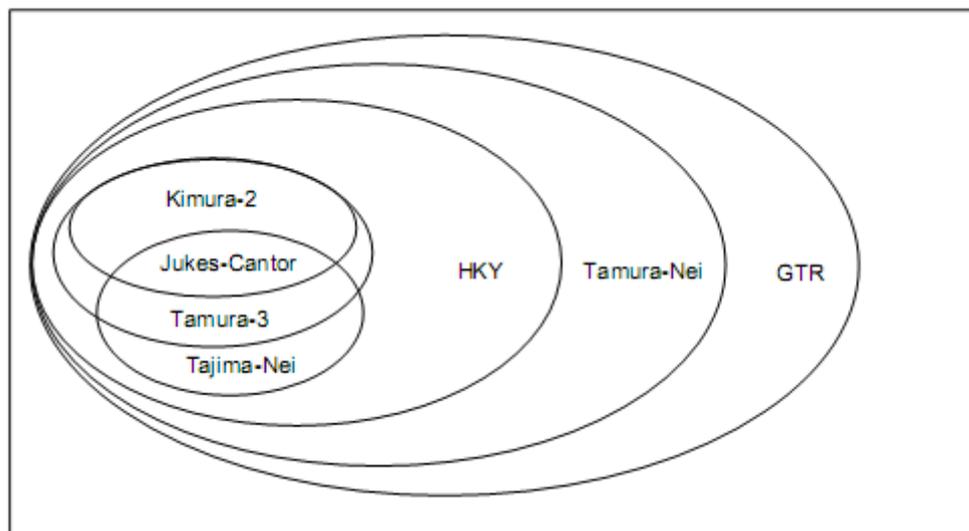


Figura 2: Relações entre os diversos modelos evolutivos, adaptado de GARCIA (2007).

Os modelos mencionados até então partem da premissa de que a taxa de substituição é constante ao longo da seqüência. Contudo, a realidade pode ser bastante distante disso. Segundo GARCIA (2007) há sítios e regiões extremamente conservados nas proteínas, indicando que seus nucleotídeos codificam partes cruciais ao funcionamento da mesma. Por outro lado, outros trechos da cadeia de aminoácidos estarão menos sujeitos a uma pressão de seleção purificadora, eventualmente por não participarem de nenhum sítio funcional ou estrutural crítico da proteína. Estes sítios estarão, portanto, mais sujeitos à fixação de mutações.

4.3 Métodos de Reconstrução de Filogenias Moleculares

Com a escolha do modelo evolutivo, o próximo passo é optar por um método de reconstrução filogenética que melhor se adeque aos dados disponíveis e ao modelo escolhido. Os métodos conhecidos para a construção de filogenias são baseados em matrizes de distâncias (a matriz de dados, nucleotídeos, por exemplo, é transformada em uma matriz de distância) ou análise de estados do caractere (nucleotídeos ou outro caractere são analisados diretamente, ou seja, sítio a sítio) (SCHNEIDER, 2003).

O método mais conhecido do primeiro grupo denomina-se *Neighbor-joining* (agrupamentos vizinhos, NJ). Já no segundo grupo, também denominados de métodos discretos, estão os métodos de Máxima Parcimônia (MP), Máxima Verossimilhança (MV).

Apesar da existência de diferentes métodos de construção de uma árvore filogenética, existem alguns princípios importantes e comuns: 1) supõe que características similares são homólogas, até que se mostre o contrário; 2) usa características derivadas compartilhadas recém herdadas; 3) relaciona características derivadas compartilhadas como marcadores de parentesco (FORATTINI 2005).

4.3.1 Máxima Parcimônia

Esse método é definido como o método que considera que uma mutação é mais provável do que duas. É um método discreto, pois atua diretamente nas seqüências ou nas funções operadas por elas.

Os métodos de máxima parcimônia buscam escolher a árvore, dentre todas as possíveis, dados os táxons da análise, que explique a evolução entre as seqüências com o menor número possível de substituições. Essa abordagem tem um forte apelo pela simplicidade do raciocínio, inspirado na “navalha de Occam”, princípio que diz que os fenômenos naturais devem ser explicados da maneira mais simples possível, com o menor número de hipóteses e parâmetros (MATIOLI, 2001).

Algumas variações do método da máxima parcimônia:

- na parcimônia generalizada, admite-se pesos diferentes entre transições e transversões por terem probabilidades diferentes de ocorrer;
- na parcimônia ponderada, dá-se maior peso a mutações que ocorram em seqüências mais informativas como, por exemplo, mutações que ocorram em regiões sabidamente conservadas;

- na variação determinada são atribuídos pesos maiores a polimorfismos que ocorram nas primeira e segunda bases codificadoras de um aminoácido, levando em conta que mutações na terceira base são, na maioria das vezes, silenciosas;
- em outra variação do método são consideradas as diferentes taxas evolutivas nas diferentes regiões do gene.

4.3.2 Métodos de Distância - NJ

Métodos de distância reduzem a variação entre duas seqüências a uma única medida de distância entre elas e trabalham com essas distâncias na estimativa da árvore final. O método de NJ é baseado em uma matriz de distância e faz uso de modelos probabilísticos de evolução.

A análise pode ser iniciada com todos os táxons ligados a um único nó central configurando o que chamamos de árvore-estrela. A cada passo é escolhido um par de nós para ser agrupado num novo nó, até que toda a árvore tenha sido resolvida. A escolha do par a ser agrupado é baseada no seguinte critério: para cada possível par de nós ainda ligados ao nó central, calcula-se qual par que, agrupado, determinaria a menor soma de comprimentos de ramo da árvore (GARCIA, 2007). Após escolhido o par que determina a menor soma total de comprimento de ramos, rearranja-se a matriz de distâncias. Todo o processo é repetido para os nós restantes, até que a árvore esteja completamente resolvida. Teremos então uma árvore não-enraizada, com comprimentos assinalados em cada ramo.

Em suma, a árvore, neste método, é estimada através do comprimento dos ramos com base nas distâncias evolutivas. O método não examina todas as topologias possíveis, mas procura encontrar vizinhos que minimizem o comprimento total da árvore, tendo bom desempenho nos casos em que a divergência entre as seqüências estudadas é baixa (HOLDER, 2003).

4.3.3 Máxima Verossimilhança

É um método discreto, assim como a Máxima Parcimônia, e busca dentre todas as árvores a mais verossímil de acordo com os dados fornecidos (seqüências do alinhamento), conforme o modelo evolutivo. Esse método permite calcular estimativas de menor variância, conseqüentemente, com menos erros de amostragem.

FELSENSTEIN (2004) propôs que a verossimilhança para cada um dos sítios seja calculada como o produto da soma das probabilidades de ocorrência de todas as substituições possíveis em um nó. Para todos os nós da árvore serão consideradas todas as possibilidades de comprimento de ramo.

Na grande maioria das análises, o espaço/tempo necessário para gerar as árvores possíveis é imensamente grande e proibitivo para uma busca exaustiva, demandando a adoção de algoritmos heurísticos de otimização de busca.

HOLDER (2003) afirma que a Máxima Verossimilhança é um método muito atrativo, porém muito pesado computacionalmente (**Tabela 1**). É uma poderosa ferramenta para análise de modelos de evolução.

TABELA COMPARATIVA			
MÉTODO	VANTAGENS	DESVANTAGENS	SOFTWARE
NJ	Rápido	Informação é perdida ao transformarmos sequências de DNA em medidas de distância	PAUP* MEGA PHYLIP
MP	Rápido para centenas de sequências; robusto se os ramos forem curtos	Pode ser inconsistente se houver variação substancial no comprimento dos ramos	PAUP* NONA MEGA PHYLIP
MV	Capta o que os dados nos dizem a respeito da filogenia em um dado modelo evolutivo	Pode ser lento demais a ponto de ser proibitivo, devido a elevada demanda computacional	PAUP* PAML PHYLIP

Tabela 1: Vantagens e desvantagens dos métodos de inferência filogenética. Adaptado de HOLDER 2003.

4.4 Árvores Filogenéticas

Árvores filogenéticas são hipóteses de relações evolucionárias (já que não se tem acesso à história) que podem ser representadas de diversas formas: cladogramas, os quais demonstram qual o ancestral mais próximo; filograma, que mostram o número de substituições que ocorreram em cada nó; e ultramétricas, que podem ser usadas para representar o tempo evolucionário propriamente dito (PAGE & HOLMES, 1998).

Assim, uma filogenia ou árvore filogenética é um gráfico que demonstra o relacionamento evolutivo entre organismos. Pode ser utilizada como ferramenta para taxonomia, análise de diversidade biológica, análise evolutiva de microrganismos e seus hospedeiros e análise epidemiológica.

O objetivo de filogenias moleculares é converter informações inicialmente dadas em seqüências de DNA em uma árvore evolucionária. A escolha da melhor forma de fazer uma análise filogenética está relacionada, principalmente, com os dados de que se dispõe. Além disso, é necessário um alinhamento confiável e escolha de modelo evolutivo apropriado.

Segundo PAGE & HOLMES (1998), árvores filogenéticas podem apresentar ou não um grupo externo como raiz ancestral, o qual dá uma direção evolucionária à árvore e diferencia os caracteres ancestrais dos derivados.

4.5 Teste de Confiança em Topologias

4.5.1 *Bootstrap*

De acordo com FELSENSTEIN (1985), o *bootstrap* é o teste mais usado para se avaliar a confiança numa filogenia molecular. Ele permite avaliar a confiança no suporte de cada nó da topologia escolhida, mediante repetição da análise filogenética sobre pseudo-réplicas do alinhamento original.

As pseudo-réplicas têm o mesmo tamanho do alinhamento original e suas colunas são sorteadas dentre as colunas do alinhamento original, com possibilidade de repetição, ou seja, uma coluna do alinhamento original pode ser sorteada mais de uma vez. O número de sorteios é igual ao número total de sítios. Desta forma, alguns sítios serão amostrados mais de uma vez enquanto outros estarão fora do pseudo-alinhamento. Para cada divergência de ramos na árvore original que exista também numa árvore réplica, soma-se um ao *bootstrap* dessa divergência. Ao final do processo divide-se o *bootstrap* de cada ramo pelo número total de réplicas analisadas e temos uma medida de confiança sobre os ramos da árvore original (MATIOLI, 2001). Geralmente, esse valor é dado em porcentagens sobre os respectivos nós das árvores filogenéticas.

5 RELÓGIO MOLECULAR

Taxas evolutivas de seqüências de DNA variam entre espécies, organelas, genes e porções do mesmo gene. No entanto, sabe-se que algumas proteínas têm taxas evolutivas constantes ao longo do tempo, como a hemoglobina, por exemplo. Essas proteínas podem ser utilizadas como cronômetros na estimativa da divergência dos táxons (MATIOLI, 2001).

Um bom exemplo seria: analisando a seqüência de DNA de determinado gene do homem e do chimpanzé, encontrássemos 3% de diferença e, sabendo-se que essas espécies divergiram há 8 milhões anos e que a diferença com roedores é de 35%, poderíamos concluir que o tempo de divergência entre humanos e roedores seria de 93 milhões anos.

Proteínas com papéis cruciais no organismo têm menor variabilidade, pois são fortemente selecionadas. Proteínas conservadas são usadas para resolver problemas antigos e proteínas variáveis são usadas para resolver problemas recentes.

Como a hipótese do relógio molecular assume que as taxas de substituições entre os ramos de uma filogenia são homogêneas, devemos verificar, portanto, antes de fazer qualquer inferência, se os táxons da filogenia em questão estão evoluindo segundo esta premissa (PAGE & HOLMES, 1998). As principais críticas ao relógio molecular estão relacionadas com o fato de que dificilmente taxas de substituições de genes comuns a espécies muito diferentes serão realmente constantes.

Para MATIOLI (2001), a constância do relógio em mutações sinônimas em mamíferos é determinada pela constância (ou não) da subjacente taxa de mutação, que é afetada pelo tempo de geração, taxa metabólica e mecanismo de reparação de DNA, conhecidos coletivamente como efeitos de linhagem. Em contraste, o relógio em mutações não sinônimas é menos sujeito a efeitos de linhagem, mas é caracterizado por episódios de substituição, que são tidos como evidências de seleção natural.

6 IMPORTÂNCIA DA ANÁLISE FILOGENÉTICA COMO FERRAMENTA EM EPIDEMIOLOGIA MOLECULAR DE DOENÇAS VIRAIS

Para FORATTINI (2005), o estudo feito sobre indivíduos de determinada população, de maneira a se observar sua composição gênica e suas relações com os agravos à saúde é a epidemiologia molecular.

Segundo HOLDER (2003), análises filogenéticas podem ser usadas em epidemiologia molecular com diversos objetivos, como por exemplo, detecção de ortologia ou paralogia e estimação de tempo de divergência em estudos evolutivos; procura por polimorfismos importantes na seleção natural, que podem aumentar a imunogenicidade de determinadas proteínas auxiliando na produção de vacinas; detecção de pontos de recombinação para funcionarem como marcadores moleculares para detecção de novas linhagens; identificação de mutações relacionadas a distúrbios ou doenças; determinação da identidade de novos patógenos e caracterização molecular.

Dados moleculares de espécies extintas são difíceis ou impossíveis de se obter; no entanto, usando características morfológicas de fósseis, encontramos uma saída para estimar as relações entre os organismos. De qualquer forma, organismos como vírus, não deixam registros fósseis e a única forma de estudar seu passado é através de relações filogenéticas de vírus existentes (VANDAMME, 2003). Para PAGE & HOLMES (1998), filogenias de vírus contem informações importantes a respeito da sua história epidemiológica e podem ser usadas para determinar onde eles se originaram e como se deu sua disseminação através das populações.

6.1 Estudos de Recombinação Genética

Como foi dito anteriormente, estudos realizados utilizando-se análises filogenéticas podem ter como finalidade um maior conhecimento a respeito das taxas evolutivas dos vírus, de que forma surgiram, se houve recombinações, detecção de ortologia ou paralogia, entre outros.

Para auxiliar no entendimento das aplicações dessa ferramenta citamos como exemplo uma nota prévia publicada por CHOI *et al.* (2002), que estudou 24 sequências de influenza A (H1N2) isoladas de suínos, nos EUA, de 1999 a 2001, de propriedades com casos clínicos da doença. Os autores tinham como objetivo caracterizar oito segmentos do genoma de cada uma das amostras virais na tentativa de rastrear sua origem.

A espécie suína possui em seu epitélio traqueal receptores para o vírus influenza com origem aviária e mamífera. Portanto, suínos são extremamente importantes no que diz respeito a recombinações entre diferentes cepas de influenza, quando sujeitos à co-infecção.

Os resultados demonstraram que as 24 amostras estudadas são recombinações entre o H1N1 clássico e a influenza suína (H3N2), reforçando a necessidade de vigilância epidemiológica.

Um estudo de HOLMES *et al.* (2005) utilizou 156 genomas completos de influenza A (H3N2) de casos ocorridos em humanos de 1999 a 2004 no estado de Nova York nos EUA. Filogenias foram geradas pra cada segmento de gene do vírus, as quais revelaram que múltiplas recombinações ocorreram. O dado mais relevante encontrado foi que no *cluster* que incluía amostras de H3N2 verificou-se que a hemaglutinina de um H3N2 coletado em 2000 foi passada por recombinação para todos os demais H3N2 amostrados de 2002 a 2003. Essa nova linhagem foi identificada como a causadora da epidemia 2002/2004. Além disso, constatou-se que diversas outras linhagens de influenza continuam co-circulando na mesma população.

Em SCHUMANN *et al.* (2009) oito isolados de *Rotavírus* do grupo A foram estudados com o intuito de investigar a existência de transmissão e recombinação de vírus interespecies. Sabe-se que os grupos de *Rotavirus* A, D, F e G são encontrados em aves; contudo, dados de sequências nessas espécies são escassos.

Os *Rotavirus* possuem 11 segmentos de RNA como material genético, sendo que cada um codifica para uma proteína, estrutural ou não. Os segmentos VP4, VP6 e VP7 foram os escolhidos para o estudo, pois são mais imunogênicos. Enquanto que o segmento NSP5 das aves foi escolhido por ter baixa similaridade em comparação com o mesmo gene das linhagens virais que afetam mamíferos.

Nas árvores filogenéticas geradas, geralmente, as amostras de aves estavam em ramos diferentes das amostras de mamíferos; no entanto, esses grupos estavam mais próximos entre si do que do *cluster* que incluía o grupo C (*Rotavirus* humano). Dentro do grupo das aves havia uma subdivisão entre as amostras provenientes de frangos (Ch) e outra com as amostras provenientes de perus e pombos (Tu e Pi). Uma amostra proveniente de frango chamada Ch2 se agrupou junto aos perus, o que representa a possibilidade de transmissão interespecífica. Uma segunda amostra de frango denominada Ch661G1 teve a sequência do fragmento VP4 não agrupada a nenhuma amostra de frango, pombo ou peru, o que pode indicar a existência de recombinação com alguma linhagem do vírus ainda não conhecida (**Figura 3**).

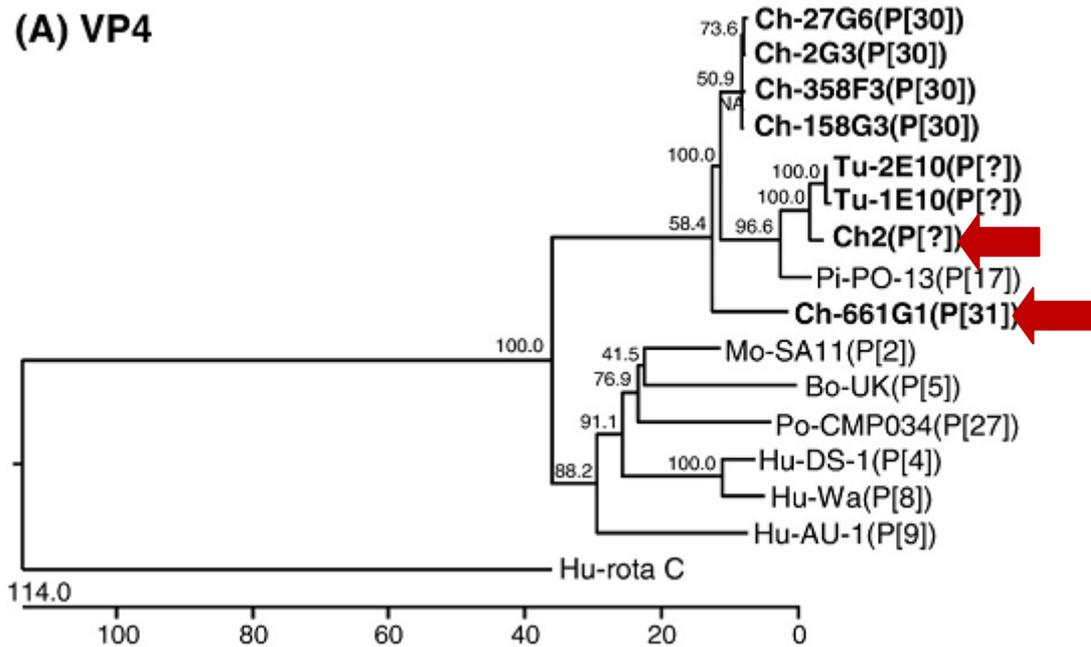


Figura 3: Árvore filogenética da região VP4 das amostras de *Rotavirus* isoladas de aves e mamíferos. Retirado de SCHUMANN *et al.* (2009).

Os autores sugerem ainda que, rotavíruses de animais podem contribuir para a variabilidade genética das rotavíruses humanas por transmissão direta ou recombinações em condições naturais.

Outro bom exemplo de aplicações de análises filogenéticas para estudos de evolução e recombinações é um trabalho, publicado no periódico *Nature*, produzido por SMITH *et al.* (2009) que investigou a origem evolucionária de amostras da influenza A (H1N1), recente causadora de pandemia. O objetivo dos autores era estimar o tempo que foi necessário para o aparecimento dessa nova linhagem de influenza, de que forma ela surgiu e quanto tempo levou para se tornar epidêmica.

Diversas sequências de influenza de diferentes espécies animais foram utilizadas no estudo, as quais foram retiradas do *NCBI Influenza Virus Resource*. Os resultados demonstraram que esse novo vírus é proveniente de cepas comuns na espécie suína. Os genes HA, NP e NS são provenientes de recombinação com influenza suína clássica, PB2 e PA derivados de influenza aviária, PB1 vindo do H3N2 humano, os quais deram origem a uma linhagem que circula há anos na América do Norte. Em contraste, os genes NA e M da nova gripe são provenientes de uma linhagem de H1N1 de aves da Eurásia (**Figura 4**). Além disso, os dados obtidos mostram que as primeiras transmissões a humanos ocorreram vários meses antes do início da epidemia e destacam a necessidade de uma vigilância sistemática de influenza em suínos, considerando que essa espécie pode servir para novas recombinações que poderiam gerar novos organismos com potencial pandêmico.

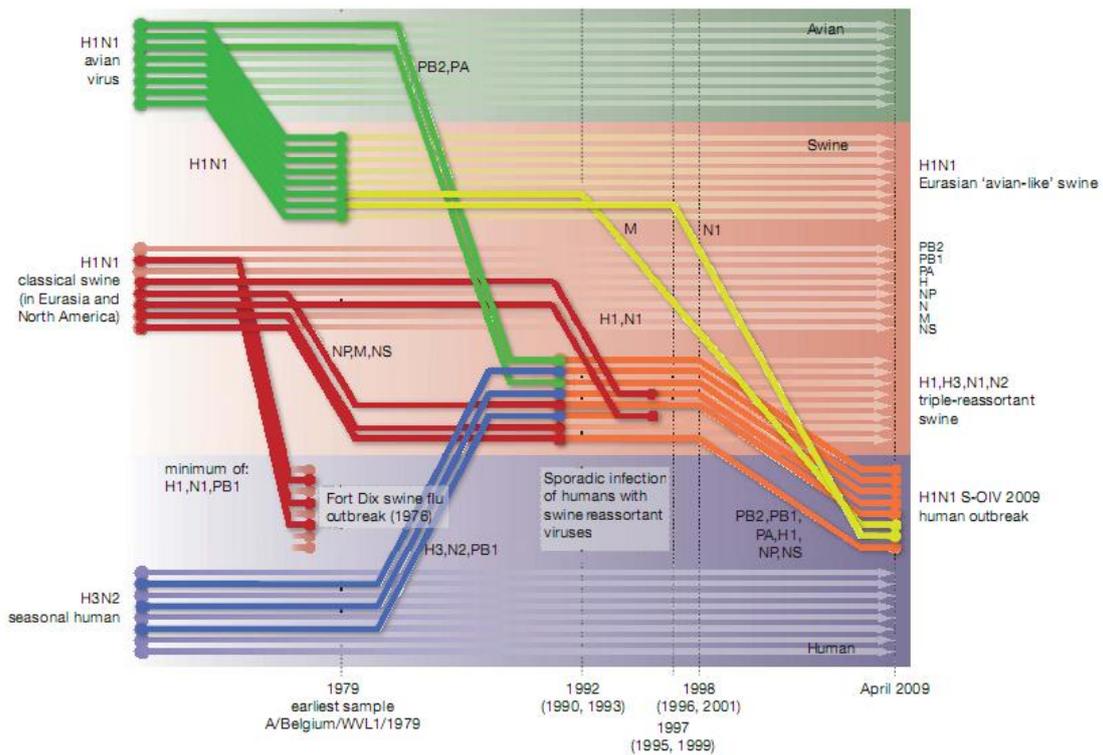


Figura 4: Reconstrução dos eventos de recombinação que levaram a emergência da recente gripe humana originada de suínos. Retirada de SMITH *et al.* (2009).

Uma hipótese gerada por estudos ecológicos afirma que linhagens de influenza, que possuam como reservatórios aves silvestres aquáticas estão em estase evolucionária; ou seja, em baixas taxas evolutivas. O estudo de CHEN (2006) avalia essa hipótese estimando taxas de substituição de nucleotídeos em diversas amostras de influenza aviária. O autor chega à conclusão de que a hipótese acima estava errada, pois as taxas de substituição de nucleotídeos encontradas eram extremamente altas, em torno de 1×10^{-3} substituições por sítio por ano. Os comprimentos dos ramos das árvores geradas confirmam essa informação, sugerindo que os vírus de aves silvestres evoluem mais rapidamente quando comparados aos de aves domésticas e que a diversidade genética encontrada também é muito grande dentro de cada subtipo e geralmente de origem recente (**Figura 5**). No entanto, os ciclos evolutivos de influenza em aves silvestres e domésticas não podem ser analisados em separados, pois a movimentação dos vírus entre as diferentes espécies de aves é constante.

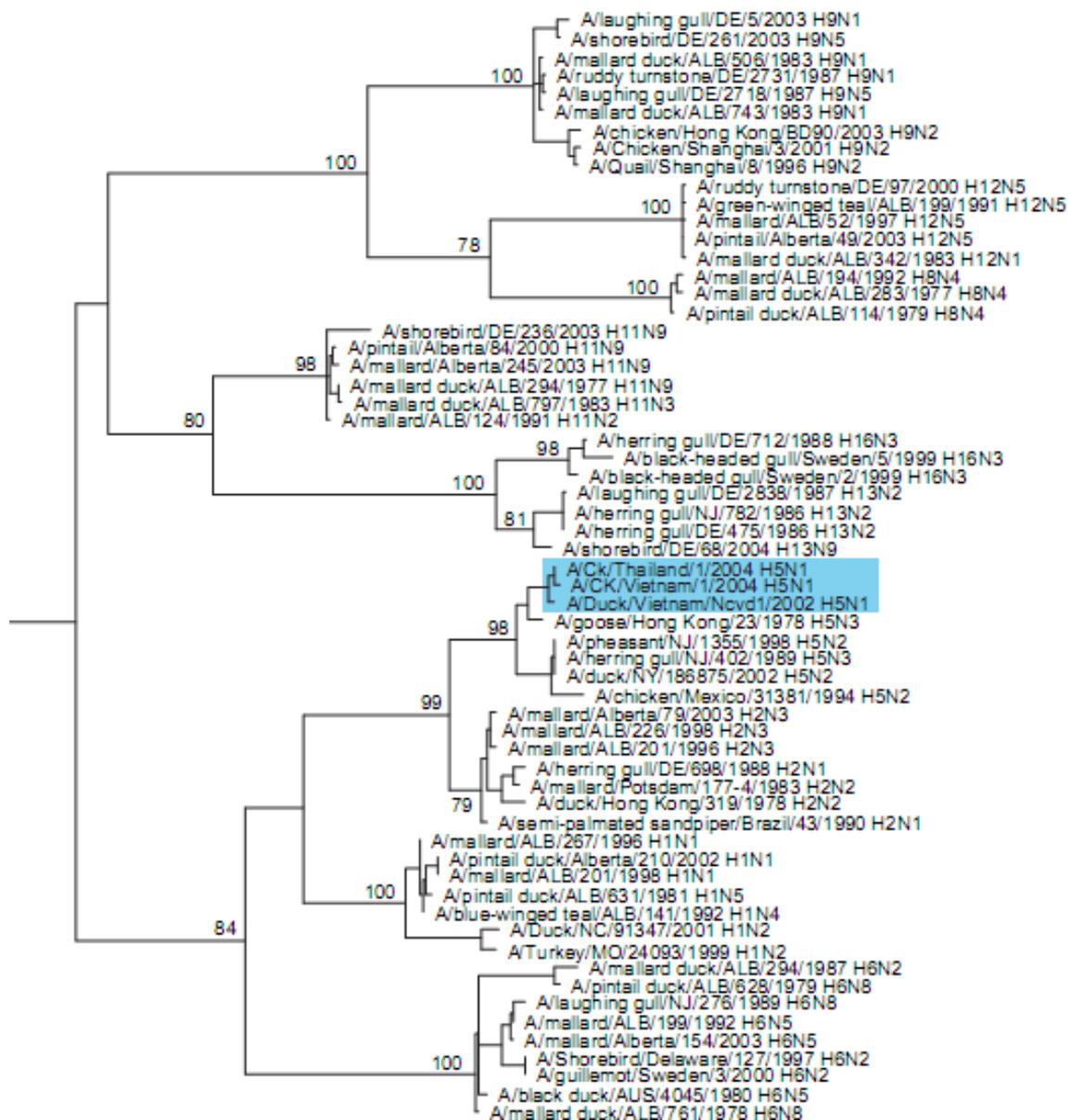


Figura 5: Árvore filogenética do gene HA. O grupo H5N1 está hachurado.

Retirado de CHEN (2006).

PISONI *et al.* (2007) descreveram a co-infecção por MVV e CAEV em duas cabras naturalmente infectadas, além de sugerirem a ocorrência de recombinação entre esses dois vírus *in vivo*, em um dos animais.

Os Lentivírus de Pequenos Ruminantes (MVV e CAEV) são estritamente relacionados quanto a características biológicas, fenotípicas e genotípicas; no entanto, nunca houve relatos de co-infecção ou recombinação entre eles. Por se tratarem de *Retrovírus*, espera-se dos mesmos alta variabilidade genética e facilidade de recombinação durante a replicação. Atualmente, a

recombinação é reconhecida como o principal fator gerador de variabilidade genética no HIV (THOMSON, 2005).

6.2 Caracterização da Diversidade Molecular

Um grande número de trabalhos científicos tem se dedicado a caracterizar, de forma molecular, novos agentes e agentes antigos com a intenção de estimar a diversidade molecular existente. No entanto, a maioria desses trabalhos traz diversos outros tipos de informações como veremos a seguir.

Uma publicação de BRANDÃO *et al.* (2006) analisou 15 sequências de *Coronavírus* bovino (BCoV) de propriedades suspeitas da doença do Sudeste brasileiro, juntamente com 10 amostras retiradas do *GenBank* de outros países como Canadá e EUA. A análise filogenética revelou a formação de dois *clusters* entre as amostras, sendo que o *cluster* I incluía as amostras dos outros países e uma amostra brasileira. No *cluster* II houve segregação das amostras portadoras de uma deleção de 18 nucleotídeos, a qual era somente encontrada em sequências de *Coronavírus* humano (HCoV-OC43), reforçando a hipótese recentemente lançada de potencial zoonótico de BCoV.

Além disso, o fragmento seqüenciado nesse estudo, região hipervariável do gene da glicoproteína espiculada, foi eleito um bom marcador genético a ser usado em maiores estudos sobre eficiência de imunógenos, pois é a proteína mais exposta à pressão seletiva imunológica, portanto, a mais propensa a polimorfismos.

KINSELLA *et al.* (2003) finalizaram o sequenciamento do genoma do Vírus da Febre Hemorrágica Criméia-Congo (CCHF) com a determinação da porção L, que codifica uma RNA polimerase RNA - dependente. Esse agente era o único vírus da categoria A de patogenicidade para humanos (mortalidade de 60%), que ainda não possuía todo o genoma disponível. Árvores filogenéticas confirmaram as informações obtidas nos alinhamentos de que os primeiros 9 a 11 nucleotídeos da região amplificada são extremamente conservados, tornando-se, assim, bons marcadores para serem utilizados em inferências de relações filogenéticas com outros grupos de vírus da mesma família, como mostra a **Figura 6**.

Mais importante do que a informação anterior, esse estudo demonstrou evidências da existência de sequências que codificam uma protease autocatalítica na região estudada.

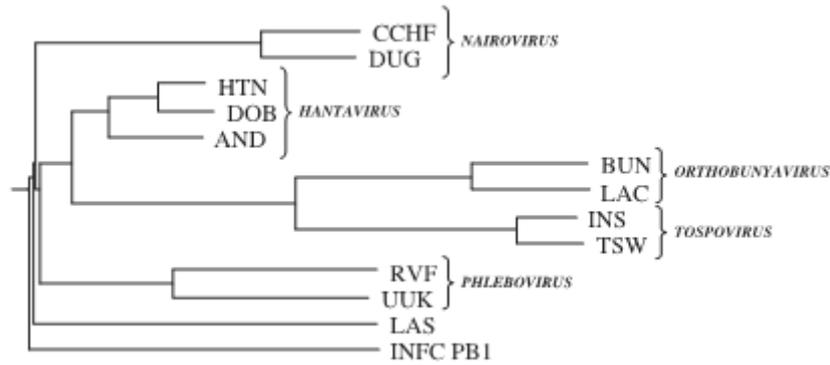


Figura 6: Árvore filogenética da Família *Bunyaviridae* baseada na sequência do fragmento L. Retirada de KINSELLA *et al.* (2003).

LIEGEOIS *et al.* (2006) caracterizaram uma nova linhagem de Vírus da Imunodeficiência Símia (SIV), o qual acomete a espécie de macacos do velho mundo *Miopithecus ogouensis*.

As amostras utilizadas são provenientes de animais de cativeiro e de vida livre. Dentre os animais analisados, dois apresentaram essa nova linhagem de vírus e análises filogenéticas confirmaram que essa é uma linhagem diferente das já conhecidas, que acometem outras espécies símias.

Vírus da Imunodeficiência são amplamente distribuídos em primatas não-humanos e tem como característica serem espécie-específicos. Em alguns grupos, análises filogenéticas sugerem co-especiação vírus/hospedeiro.

HU *et al.* (2001) analisaram o gene que codifica uma proteína de superfície do Vírus da Hepatite B de Chimpanzés (ChHBV). Os animais estudados eram cronicamente infectados, o que ocorre por infecção ao nascimento. Todos haviam nascido em cativeiro ou foram capturados após a proibição da caça de chimpanzés. Segundo a literatura, cada grupo de primatas parece ser infectado por uma linhagem específica de vírus. Na análise filogenética dos vírus, os pesquisadores observaram a formação de três *clusters*. Para verificar se os chimpanzés estudados pertenciam a subespécies diferentes e, por isso, os vírus segregariam em *clusters* diferentes, o DNAm desses mesmos animais foi analisado a partir de bulbo piloso. Os resultados mostraram que os animais pertenciam a subespécies diferentes, de regiões geograficamente separadas no Oeste africano.

Roedores capturados na região transdanubiana da Hungria foram investigados por PCR em tempo real para detecção e caracterização molecular de Dobrava *Hantavírus* (DOBV) em estudo realizado por JAKAB *et al.* (2007). Dos 22 animais capturados, três foram positivos. Na análise

filogenética, observou-se que as amostras húngaras eram muito similares a uma da Eslovênia (94 a 97% de homologia) (**Figura 7**), sendo que uma delas (EF59980) era proveniente de animal capturado próximo à fronteira com a Eslovênia, um dos locais onde o vírus foi descrito pela primeira vez.

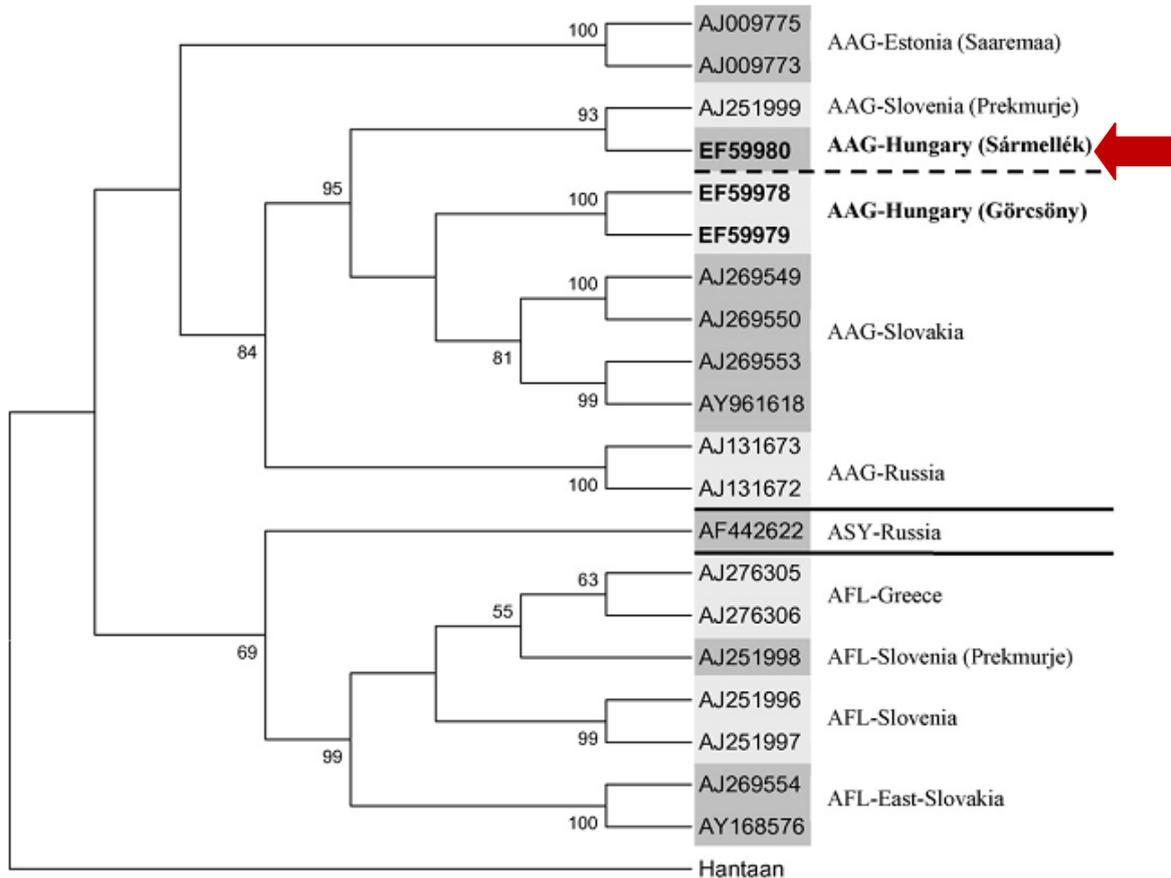


Figura 7: Árvore publicada em JAKAB et al. (2007) demonstrando a proximidade filogenética das amostras de Hantavírus eslovena e húngaras.

Em um trabalho de PISONI *et al.* (2005), foi realizada análise filogenética com o principal objetivo de estabelecer uma filogenia de SRLV para servir como base para futuros estudos de epidemiologia molecular de Lentivírus de Pequenos Ruminantes em rebanhos de caprinos leiteiros do Norte da Itália, com interesse no monitoramento dos efeitos do avanço da doença e futuros programas de erradicação da mesma.

A região do genoma analisada foi um pequeno fragmento dentro do gene *gag*, MA o qual codifica a proteína matriz; 52 novas sequências foram estabelecidas (**Figura 8**).

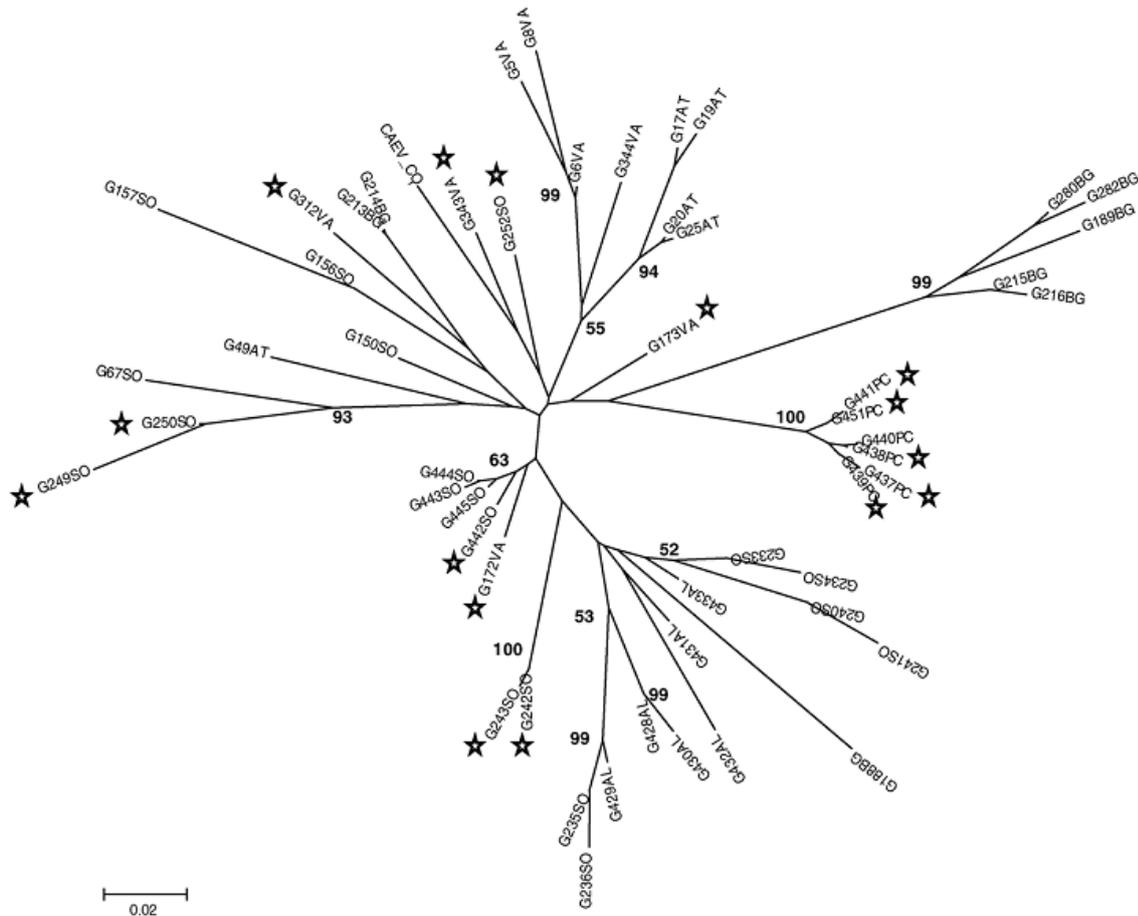


Figura 8: Análise filogenética incluindo todas as amostras de SRLV estudadas. As estrelas identificam amostras de animais com sinais clínicos. Retirado de PISONI *et al.* (2005).

Os resultados obtidos demonstraram que as amostras italianas segregaram em grupos muito próximos e que são similares ao protótipo CAEV-CO (**Figura 9**). Isso sugere uma origem comum ao SRLV encontrado nos rebanhos monitorados. Possivelmente relacionado à introdução de animais positivos em rebanhos negativos.

Além disso, as amostras analisadas apresentaram alta variabilidade no fragmento estudado e uma inserção de 7 aminoácidos em todas as amostras de MVV, o que possibilita a diferenciação de amostras de CAEV. Os autores concluíram que o fragmento MA é adequado a estudos filogenéticos e pode ser aplicado em programas de monitoramento e erradicação.

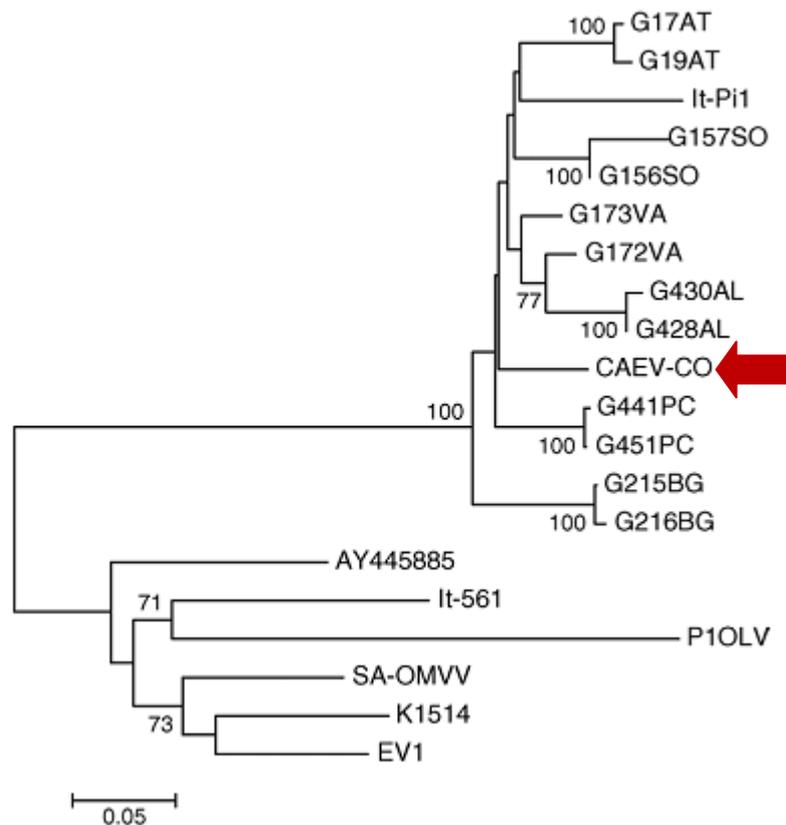


Figura 9: Análise filogenética de algumas amostras italianas segregando junto ao protótipo CAEV-CO (EUA). Retirada de PISONI *et al.* (2005).

AAMIR *et al.* (2006) realizaram um estudo de caracterização molecular objetivando estabelecer as relações filogenéticas existentes entre isolados de influenza aviária (H2N9) do Oriente Médio, outros isolados asiáticos e sete novas amostras dos Emirados Árabes Unidos coletadas de 2000 a 2003. Após a epidemia causada por H5N1 os autores consideraram importante a caracterização de outras linhagens virais para investigação quanto ao seu potencial pandêmico.

Para a inferência filogenética, oito segmentos de gene foram analisados. Das sete novas amostras estudadas por AAMIR *et al.* (2006) todas tiveram a adição de um aminoácido na hemaglutinina (HA), seis continham uma mutação associada ao aumento da afinidade com os receptores de células humanas. Os genes que codificam para as glicoproteínas de superfície dos vírus e para a maioria das proteínas internas tiveram 90% de similaridade com uma linhagem de Hong Kong que causou mortes em crianças em 1997. O gene que codifica a neuraminidase teve quatro substituições, como aconteceu com os vírus pandêmicos de humanos. A sequência do fragmento M2 mostrou a troca de dois aminoácidos, o que está associado ao aumento de resistência à amantadina

em cultivo celular. Devido às características supracitadas é sugerido pelos autores que H2N9 seja potencialmente pandêmico.

KNOWLES *et al.* (1998) realizou estudo com uma sequência parcial de nucleotídeos do Vírus da Encefalomiocardite (EMCV), o qual causou surtos de miocardite fatal em suínos na Itália. A transmissão desse vírus aos suínos é, provavelmente, feita por roedores silvestres. Seis sequências de suínos e 3 de roedores silvestres da Itália foram utilizadas em inferência filogenética para verificar a existência de relação entre elas. Além disso, amostras de outros países, como Grécia, Bélgica e Países Baixos, foram inseridas na árvore filogenética.

Todas as mostras italianas estão proximamente relacionadas (94,6% de identidade), mesmo tendo sido isoladas de regiões distintas dentro desse país, o que sustenta a hipótese da transmissão aos porcos pelos roedores. Quando comparadas à maioria das amostras de outros países da Europa, as amostras foram divergentes; no entanto, constatou-se de 95,3 a 99,3% de identidade com isolados dos Países Baixos (**Figura 10**), o que sugere movimento do vírus entre regiões geográficas distantes. O vírus da febre aftosa também pode causar miocardite aguda, antes das clássicas lesões vesiculares, portanto a confirmação do diagnóstico de EMC deve ser contínua, para fins de diagnóstico diferencial com febre aftosa.

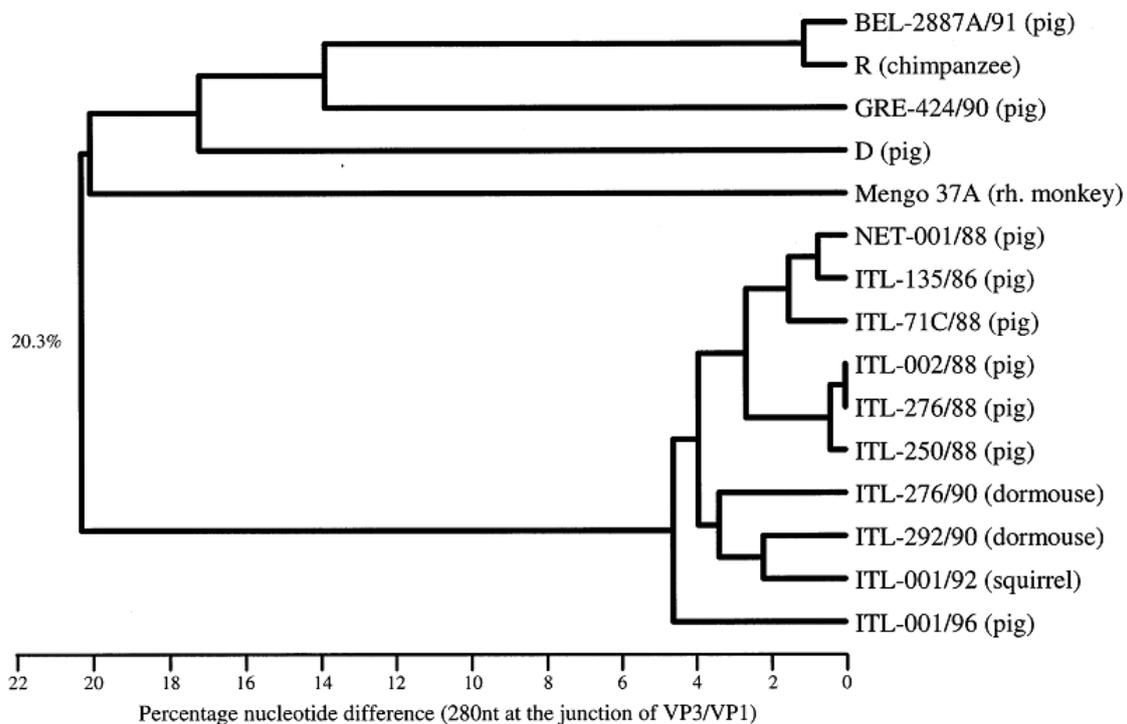


Figura 10: Árvore filogenética retirada de KNOWLES *et al.* (1998) demonstrando a similaridade entre amostras italianas e amostra dos Países Baixos.

A filogenia de 14 amostras de *Flavivirus* brasileiras foi estudada por BATISTA *et al.* (2000). Os fragmentos utilizados compreendiam uma porção do gene NS5 até uma região não codificante 3', a qual possui regiões conservadas e hipervariáveis na mesma proporção, tornando-se uma fonte rica de informações filogenéticas. Além das amostras brasileiras, cinco sequências do *GenBank* foram inseridas na árvore para dar maiores informações a respeito de divergência com outros *Flavivirus*.

Baseando-se na filogenia, a maior parte das amostras brasileiras se agruparam em dois ramos, o primeiro da febre-amarela (YF) e o segundo dividido em dengue subtipos 1, 2 e 4. Há um ramo onde houve segregação: JEV (*Japanese Encephalites Virus Complex*), (*Saint Louis Encephalites*), Ilhéus, Rocio e Cacipacoré. O vírus Iguape aparece como um separado e distante ramo (**Figura 11**).

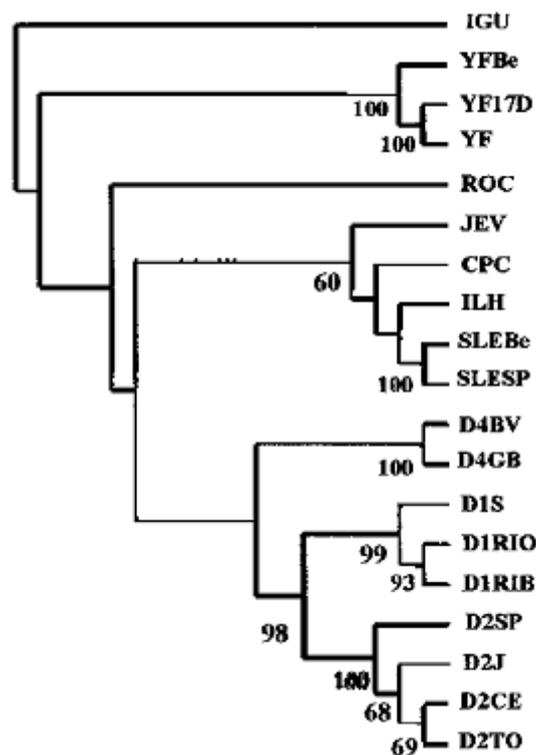


Figura 11: Árvore filogenética do grupo dos *Flavivirus*, retirada de BATISTA *et al.* (2000).

Os dados mostrados em BATISTA *et al.* (2000) concordam com os de KUNO *et al.* (1998). Os dois trabalhos são importantes para o entendimento da evolução do grupo dos *Flavivirus*.

Simian Foamy Virus (SFV) é um *Retrovirus* exógeno comum em primatas não-humanos que oferece risco ocupacional a pessoas que trabalham em zoológicos e afins. Aparentemente, não há sinais clínicos nessa doença, sendo que a infecção é detectada através de diagnósticos sorológico; no

entanto, é necessária a avaliação de um maior número de indivíduos infectados para se ter certeza quanto a essas informações. HUSSAIN *et al.* (2002) avaliaram a viabilidade de uso do método de *Western Blot* para diagnóstico de SFV em primatas não-humanos de vida livre com o intuito de disponibilizarem dados a respeito da prevalência de SFV nesses animais. Nesse estudo, diversos gêneros de primatas nunca descritos como hospedeiros desse vírus foram positivos; portanto, foram submetidos ao diagnóstico molecular e posterior reconstrução filogenética.

A análise filogenética resultou na formação de três grupos correspondentes a três linhagens diferentes do vírus, nunca publicadas e altamente divergentes, conforme **Figura 12**. As quatro novas linhagens descritas são as de *Allenopithecus*, *Trachypithecus* e *Hylobates*.

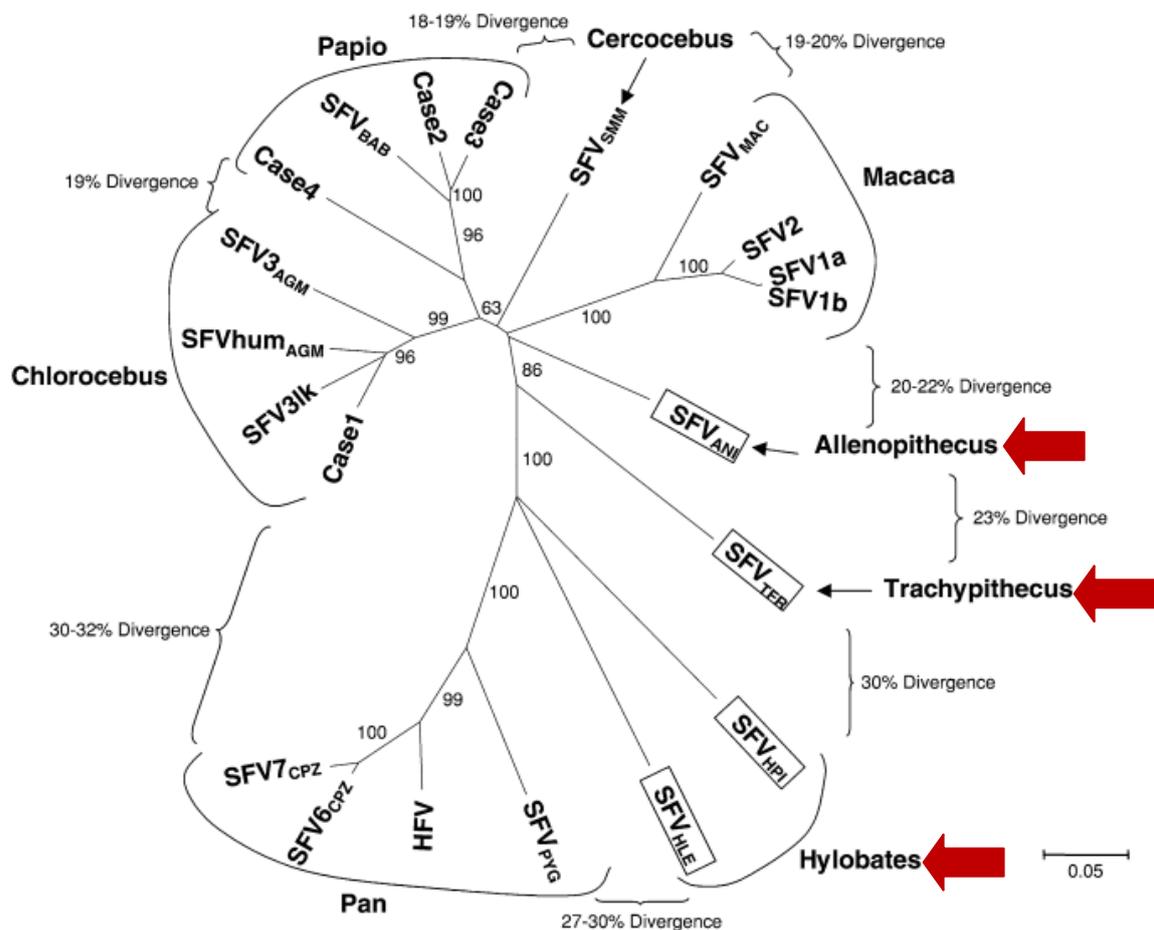


Figura 12: Reconstrução filogenética contendo sequências de SFV provenientes de diversos grupos de primatas não-humanos. Retirada de HUSSAIN *et al.* (2002).

MONCAYO *et al.* (2000), caracterizaram antigênica e geneticamente pela primeira vez um novo *Arenavirus*, patógeno de ratos da região amazônica do Peru, o *Allpahuayo Virus*. Os genes NP

e GPC foram seqüenciados e o novo vírus se enquadrou no clado/*cluster* A (**Figura 13**), que corresponde à linhagem A do Complexo Tacaribe dos *Arenavírus* do Novo Mundo, enquanto que os *Arenavírus* que causam febre hemorrágica em humanos estão no clado B.

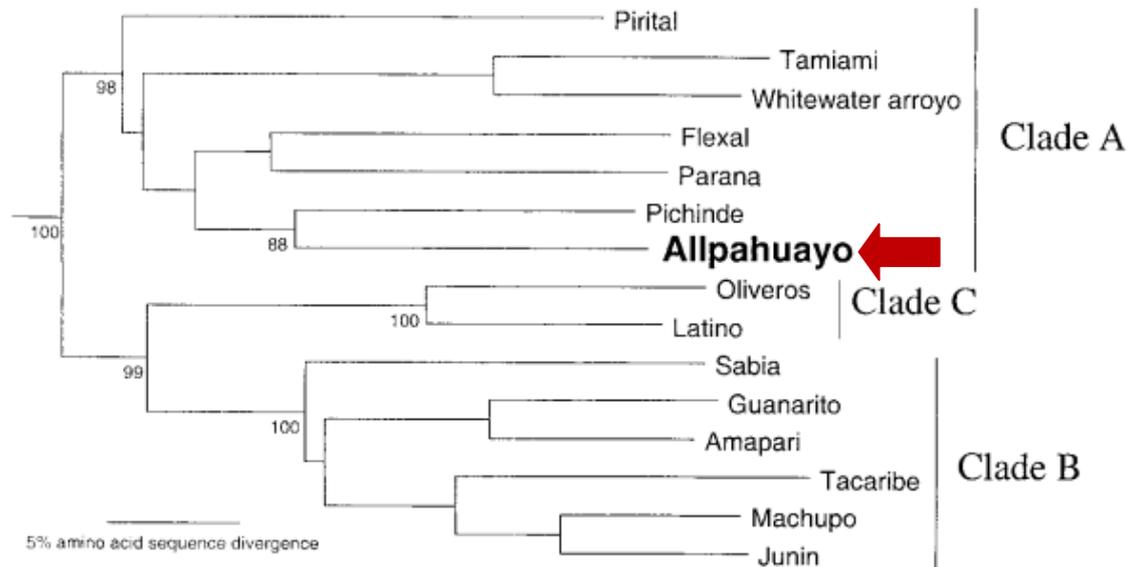


Figura 13: Árvore filogenética dos *Arenavírus*. Retirada de MONCAYO *et al.* (2000).

LECOMPTE *et al.* (2007) caracterizou um novo *Arenavírus* fortemente relacionado com o Virus da Coriomeningite Lifocítica (LCMV) pertencente aos *Arenavírus* do Velho Mundo. Essa nova linhagem foi chamada Kodoko e foi encontrada em região geográfica e hospedeiro diferentes dos comuns ao LCMV.

Cinquenta e quatro amostras de raiva provenientes de casos de raiva, transmitida por morcegos hematófagos ou relacionados com esses animais, foram caracterizadas na Argentina por CISTERNA *et al.* (2004).

Identificou-se a circulação de quatro variantes antigênicas. Através da análise filogenética verificou-se que duas variantes circulam independentemente no ciclo urbano. Com os dados, os autores puderam identificar as linhagens atuantes em cada região do país, inclusive em regiões consideradas livres da doença. Foi o primeiro trabalho de nível nacional de epidemiologia de raiva e o que se espera é que auxilie nos programas de vigilância, reduzindo o número de casos de raiva humana.

JUNIOR *et al.* (2006) estudou o gene da nucleoproteína do vírus da raiva de 20 amostras provenientes de animais domésticos (cães e gatos) e canídeos silvestres do Nordeste do Brasil.

Geralmente, amostras de cães e gatos pertencem a variante antigênica 2, e foi o que aconteceu nesse trabalho. No entanto, dentre as amostras de canídeos silvestres, três se enquadraram na variação antigênica 2, enquanto que as outras nove apresentaram padrão antigênico desconhecido.

Na análise filogenética percebemos claramente a presença de dois distintos ciclos de raiva ocorrendo entre canídeos no Nordeste do Brasil: uma relacionada a canídeos silvestres, outra relacionada aos domésticos. Conforme **Figura 14**, os sub-*clusters* 1-A e 1-B tem 95% de identidade sugerindo que as linhagens circulantes em canídeos domésticos são muito similares geneticamente.

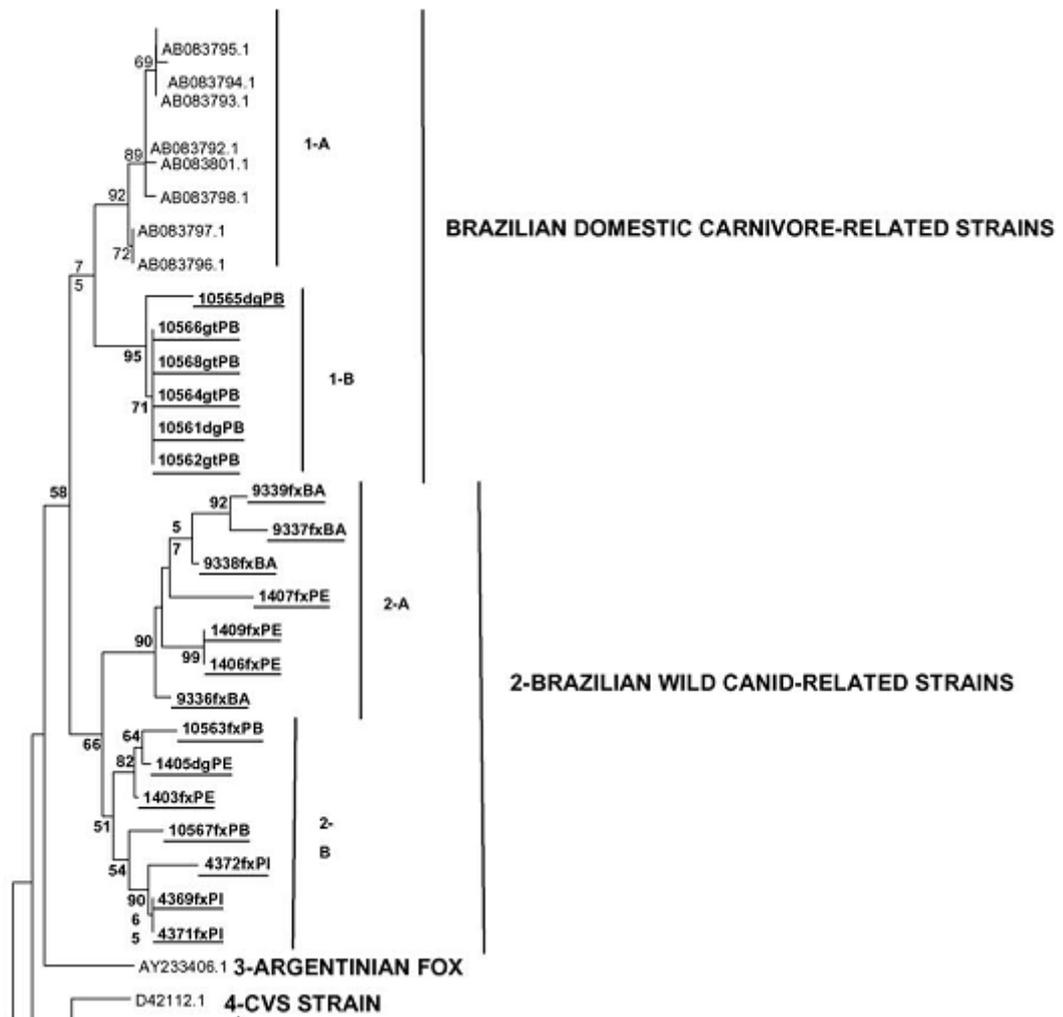


Figura 14: Árvore filogenética representativa das linhagens do vírus da raiva circulantes em canídeos silvestres e domésticos no Nordeste do Brasil. Retirada de JUNIOR *et al.* (2006).

Já nos sub-clusters 2-A e 2-B a identidade genética entre as amostras é bem menor, o que pode ocorrer devido às distâncias geográficas que separam os animais silvestres, reduzindo o contato entre os mesmos, ou devido ao menor número de animais ou, ainda, à resistência a determinadas

cepas virais. Além disso, a amostra 1405dgPB, de cão doméstico, segregou junto ao *cluster 2*, sugerindo a transmissão da raiva silvestre para a urbana, o que é raramente descrito.

VILLARREAL *et al.* (2007) caracterizou amostras do Vírus da Bronquite Infecciosa das Galinhas isolados de conteúdo entérico de aves de estabelecimentos de produção brasileiros. Mais de 20 sorotipos do vírus são conhecidos e essa diversidade de sorotipos pode ser responsável pelas falhas vacinais e, conseqüentemente, novas epidemias. No Brasil apenas uma vacina é aceita pelo Ministério da Agricultura, a do sorotipo Massachusetts; no entanto, diversos outros sorotipos são encontrados no país. Através da caracterização molecular, os autores pretendem provar a circulação de outros sorotipos no país e demonstrar a necessidade de uso de vacinas com outras cepas.

A análise filogenética mostrou que as amostras segregaram junto ao sorotipo D274 e distante do sorotipo Massachusetts, demonstrando os prováveis motivos da ineficiência do programa de controle da doença no Brasil.

6.3 Estudos de Dinâmica Populacional

A estrutura genética da população de patógenos humanos quase sempre reflete padrões conhecidos de migração humana. Segundo REAL & BIEK (2007), no caso de populações animais, parasitas transmitidos de forma direta, algumas vezes, fornecem informações substanciais a respeito de características temporais e espaciais do contato hospedeiro-hospedeiro. Por isso, estudos que buscam informações a respeito de estrutura populacional, história demográfica, entre outros, utilizam análises filogenéticas de parasitas para rastrear informações sobre seus hospedeiros.

Um estudo publicado por ANTUNES *et al.* em 2008 utilizou marcadores moleculares (microssatélites) do leão (*Panthera Leo*) e do FIV para fazer a caracterização da variabilidade genética, traçar a história demográfica dos leões e discutir aspectos da conservação dessa espécie animal.

As análises filogenéticas mostraram a existência de seis subtipos circulantes (**Figura 15**) contrariando a hipótese de que os leões africanos são provenientes de uma única população panmítica. Os subtipos de FIV estão distribuídos geograficamente em três padrões principais, devido a características de que esses leões vivem em bandos consideravelmente distantes uns dos outros.

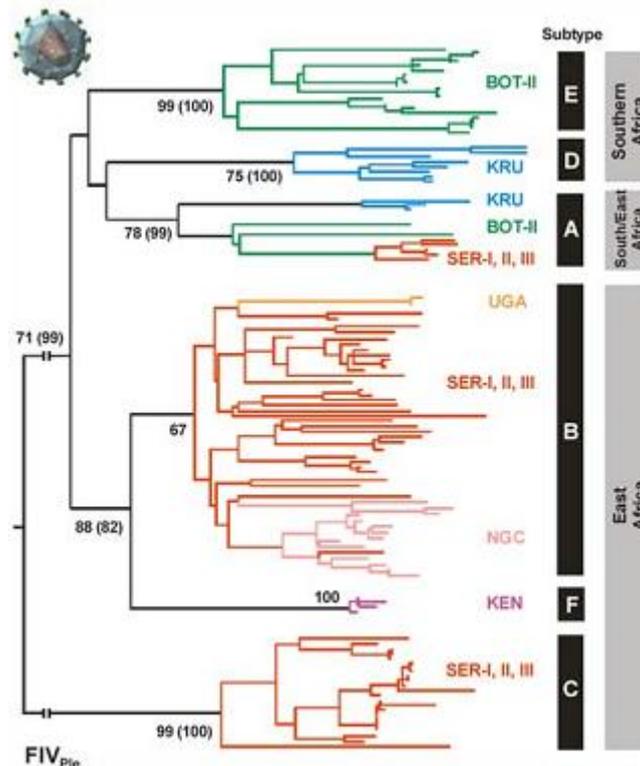


Figura 15: Os diferentes subtipos de FIV são demonstrados nessa árvore. Retirada de ANTUNES *et al.* (2008).

Estudo publicado por BIEK (2006) utiliza a rápida evolução do Vírus da Imunodeficiência Felina para revelar detalhes de estrutura populacional recente de *Puma concolor* (puma). Pumas são animais com alta prevalência de infecção por FIV sem sofrerem consequências graves por isso. A transmissão poder ser horizontal ou vertical, exigindo contato direto e causando infecções crônicas.

Mais de trezentos animais de duas diferentes regiões (A e B) foram estudados, sendo que 25% deles estavam contaminados com o vírus, sete foram encontrados com co-infecção dos diferentes subtipos de FIV, o que pode levar a existência de recombinantes. As análises filogenéticas geradas revelaram a divisão das amostras em dois *clusters* principais e oito linhagens (**Figura 16**). Os *clusters* correspondiam às duas regiões distintas estudadas, em uma delas foi encontrada menor variabilidade que deve ter ocorrido devido a uma redução na população de pumas ocorrida por aumento intensivo na caça desses animais nessa localidade.

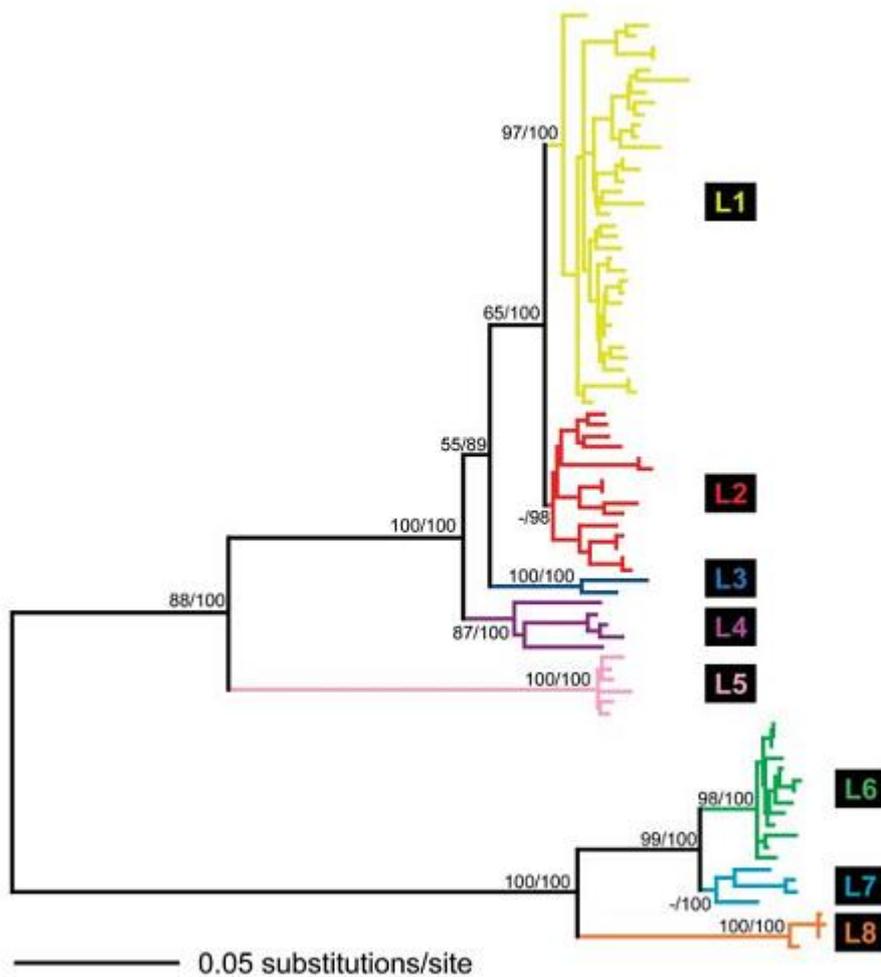


Figura 16: Árvore filogenética construída com as sequências dos genes *env* e *pol* que mostra a divisão das linhagens virais em oito subgrupos com 5% de divergência entre eles. Retirada de BIEK (2006).

6.4 Outros

Há diversas outras formas de utilização de análises filogenéticas. PISONI *et al.* (2005) demonstrou a transmissão natural de CAEV subtipo B1 de caprinos para ovinos. A transmissão ocorreu a campo em três diferentes fazendas na Itália, onde houve mistura de rebanhos caprinos e ovinos no mesmo piquete.

A inferência filogenética mostrou a formação de três diferentes *clusters* envolvendo as amostras estudadas, separando-as de acordo com a propriedade de origem dos animais (**Figura 17**).

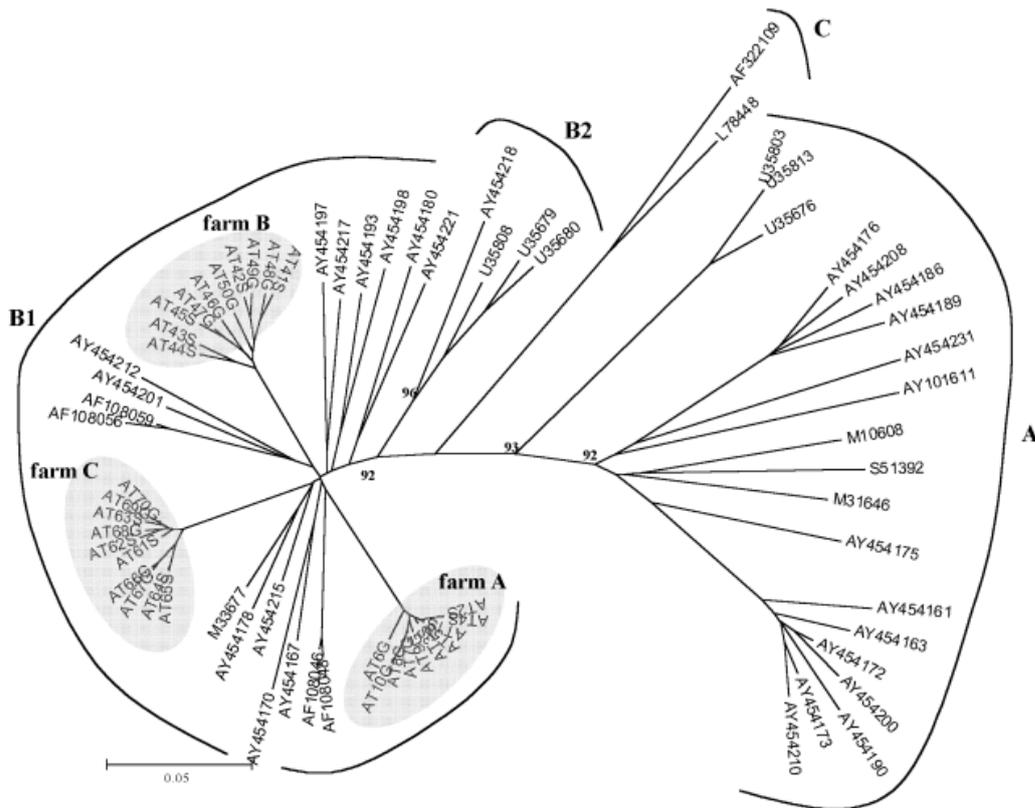


Figura 17: Árvore não enraizada representando as relações entre SRLV e seus subtipos, incluindo amostras subtipo B1 de fazendas italianas onde houve mistura de rebanhos ovinos e caprinos. Retirada de PISONI *et al.* (2005).

Os resultados obtidos através desse estudo nos fazem pensar que programas de erradicação de SRLV não devem ser voltados somente para caprinos ou ovinos, e sim para as duas espécies animais.

Um segundo trabalho de PISONI *et al.*, em 2007, utilizou análise filogenética para demonstrar a compartimentalização dos diferentes subtipos de SRLV no sangue periférico e no colostro de caprinos. O DNA proviral presente no sangue e no colostro foram analisados. Do colostro também foi analisado RNA, que corresponde aos vírus livres presentes no mesmo.

Quatro animais foram estudados e em três desses animais houve a compartimentalização dos subtipos nos diferentes tecidos. Essa compartimentalização sugere que a infecção lactogênica envolve provirus específicos não presentes na população proviral encontrada no sangue periférico.

Surpreendentemente, as sequências encontradas em amostras de sangue não são sempre as mesmas encontradas em vírus livre de amostras de colostro, o que pode ocorrer devido ao fato dos diferentes tipos celulares atuarem como reservatórios para o vírus, onde o mesmo se replica, recombina e sofre mutações de acordo com tipo celular específico dando origem a quasispécies.

Ao estudar epidemias de doenças infecciosas em felídeos silvestres, detalhes não usuais a respeito da forma que essas populações têm sobrevivido e sobre o surgimento de resistência genética nesses animais tem se revelado. Em um trabalho publicado por O'BRIEN *et al.* (2006) uma epidemia de *Coronavírus* altamente virulento ocorrida em *cheetahs* africanas e a situação endêmica de FIV em felídeos selvagens foram estudadas. O objetivo era gerar informação útil a respeito de evolução a partir da interação patógeno-hospedeiro a serem aplicadas para SARS e HIV humanos.

As análises filogenéticas geradas nesse estudo auxiliaram no conhecimento da história da população e seus hospedeiros, o que possibilitou a documentação de padrões de adaptação e co-evolução. Como exemplo, temos a árvore a seguir (**Figura 18**) que mostra a linhagem I, a qual pertence às coronaviroses de felídeos (FIPV), a gastroenterite transmissível de suínos (TGEV) e o *Coronavírus* canino (CcoV). A FIPV, em uma reserva que abrigava animais de vida livre, matou 60% das *cheetahs*, enquanto que em felinos domésticos a mortalidade está em torno de 5%. Conforme a árvore filogenética, as linhagens virais provenientes de felídeos silvestres (Aju) e domésticos (FIP) estão agrupadas de forma muito próxima, o que nos leva a crer que a explicação a respeito da diferença de mortalidade ocorrida está na genética do hospedeiro e não do parasito.

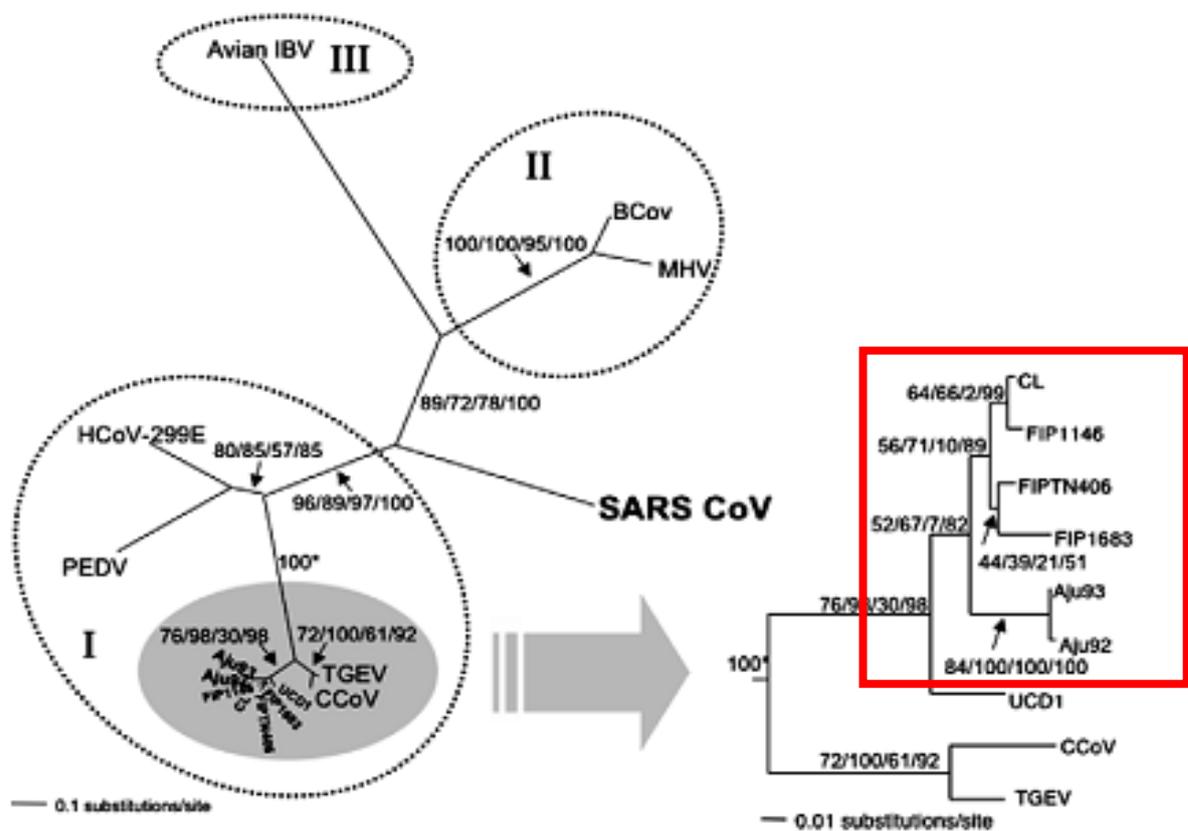


Figura 18: Árvores representando as relações filogenéticas entre *Coronavírus*. Retirada

de O'BRIEN *et al.* (2006).

Além disso, o estudo demonstra a necessidade de se utilizar ferramentas de biologia molecular disponíveis para investigações epidemiológicas de doenças humanas, assim como para a conservação de espécies animais ameaçadas de extinção.

7 CONSIDERAÇÕES FINAIS

A seqüência de DNA não é só um registro de relações filogenéticas e tempos de divergência, mas também guarda marcas de como o processo evolucionário deu forma a sua história e também o tempo decorrido no processo evolutivo das populações (PAGE & HOLMES, 1998).

Biologistas estão expandindo o uso de filogenias para um número muito grande de finalidades e os virologistas estão traçando a diversificação de linhagens virais em diferentes classes, famílias, gêneros, espécies e dentro de um único indivíduo. É importante ressaltar que para os diferentes dados disponíveis e diferentes finalidades das filogenias, os métodos de reconstrução e modelos evolutivos devem ser escolhidos de forma criteriosa.

Na veterinária, a utilização desta ferramenta auxilia na elaboração de programas de controle e erradicação de doenças, como é o caso dos Lentivírus de Pequenos Ruminantes, por exemplo. É importante para rastrear os locais ou espécies animais de onde comumente surgem as novas linhagens virais, como vimos nas epidemias de Influenza. Deve ser usadas para estudos de dinâmica populacional, possibilitando relacionar variabilidade genética viral com variabilidade genética do hospedeiro. A filogenia para caracterização molecular, estabelecendo o conhecimento das linhagens circulantes em determinados locais e, assim, possibilitando elaborar estratégias de imunização condizentes com as necessidades reais.

Concluindo, as análises de seqüências genômicas têm demonstrado um amplo espectro de aplicações, desde a classificação de microrganismos até o histórico de sua relação com o hospedeiro.

REFERÊNCIAS

- AAMIR, U.B. *et al.* Characterization of avian H9N2 influenza viruses from United Arab Emirates 200 to 2003. **Virology**, San Diego, v. 361, p. 45-55, 2007.
- ABASCAL, F., ZARDOYA R., POSADA, D. ProtTest: selection of best-fit models of protein evolution. **Bioinformatics** v. 21, p. 2104–2105, 2005.
- ANDRÉS, D. *et al.* Diagnostic tests for small ruminant lentiviruses. **Veterinary Microbiology**, Amsterdam, v. 107, p. 49-62, 2005.
- ANTUNES, A. *et al.* The evolutionary dynamics of the lion (*Panthera leo*) revealed by host and viral population genomics. **Public Library of Science**, San Francisco, v. 4, p. 1-9, 2008.
- BATISTA, W.C. *et al.* Phylogenetic analysis of Brazilian Flavivirus using nucleotide sequences of parts of NS5 gene and 3' non-coding regions. **Virus Research**, Amsterdam, v. 75, p. 35-42, 2001.
- BIEK, R.; DRUMMOND, A.J.; POSS, M. A virus reveals population structure and recent demographic history of its carnivore host. **Science**, Washington, v. 311, p. 538-541, 2006.
- BRANDÃO, P.E. *et al.* A coronavirus detected in the vampire bat *Desmodus rotundus*. **Brazilian Journal of Infectious Diseases**, Salvador, v. 12, p. 466-468, 2008.
- BRANDÃO, P.E. *et al.* On the etiology of an outbreak of Winter dysentery in dairy cows in Brazil. **Pesquisa Veterinária Brasileira**, Rio de Janeiro, v. 27, p. 398-402, 2007.
- BRANDÃO, P.E. *et al.* Molecular analysis of Brazilian strains of bovine *Coronavirus* (BCoV) reveals a deletion within the hypervariable region of the S1 subunit of the spike glycoprotein also found in human *Coronavirus* OC43. **Archives of Virology**, New York, v. 151, p. 1735-1748, 2006.
- BROWN, T.A. **DNA sequencing – the basics**. Oxford: Basics, 1994.
- BULT, C. Complete genome sequence of the Methanogenic Archaeon. **Science**, Washington, v.273, n. 5278, p. 1058-1073, 1996.

- CHEN, R.; HOLMES, E.C. Avian Influenza Virus exhibits rapid evolutionary dynamics. **Molecular Biology and Evolution**, Oxford, v. 23, p. 2336-2341, 2006.
- CHOI, Y.K. *et al.* Phylogenetic analysis of H1N2 isolates of Influenza A Virus from pigs in the United States. **Virus Research**, Amsterdam, v. 87, p. 173-179, 2002.
- CISTERNA, D. Antigenic and molecular characterization of rabies virus in Argentina. **Virus Research**, Amsterdam, v.109, p. 139-147, 2005.
- DE ROBERTIS, E.M.F.; HIB, J. **Bases da biologia celular e molecular**. Rio de Janeiro: Guanabara Koogan, 2006.
- DORMAN, J.S. Molecular epidemiology: The impact of molecular biology in epidemiology research. Disponível em: <http://www.scielo.cl/scielo.php?pid=S0034-98872000001100012&script=sci_arttext>. Acesso em 20 out 2009.
- DUTRA, C.G. **Filogeografia de *Passiflora ovalis***.2008. 45 f. Trabalho de conclusão de curso. Faculdade de Ciências Biológicas, Universidade Federal do Rio Grande do Sul, Porto Alegre/RS.
- ÉTIENNE, J.; MILLOT, F.; CERQUEIRA, A.J. **Bioquímica genética e biologia molecular**. São Paulo: Santos, 2003.
- EWING, B.; GREEN, P. Base-calling of automated sequencer traces usingPhred. II. Error probabilities. **Genome Research**, Cold Spring Harbor, v. 8, p. 186-194, 1998.
- EWING, B.; HILLIER, L.; WENDL, M.C. Base-calling of automated sequencer traces usingPhred. I. Accuracy assessment. **Genome Research**, Cold Spring Harbor, v. 8, p. 175-185, 1998.
- FELSENSTEIN, J. **Confidence limits on Phylogenies: An Approach Using the Bootstrap**. **Evolution**, v. 39, n. 4, p. 783-791, 1985.
- FERREIRA, H.B.; PASSAGLIA, L.M.P.; ZAHA, A. **Biologia molecular básica**. Porto Alegre: Mercado Aberto, 2003.
- FORATTINI, O.P. **Conceitos básicos de epidemiologia molecular**. São Paulo: Universidade de São Paulo.2004.

GARCIA, M. **Uma filogenia mitocondrial de metazoários**. 2007, 212 f. Dissertação (Mestrado em Modelagem Computacional). Laboratório Nacional de Computação Científica, Petrópolis/RJ.

HOLDER, M.; LEWIS, P.O. Phylogeny estimation: traditional and Bayesian approaches. **Nature**, London, v. 4, p. 275-284, 2003.

HOLMES, E.C. *et al.* Whole-genome analysis of Human Influenza A Virus reveals multiple persistent lineages and reassortment among recent H3N2 Virus. **Public Library of Science**, San Francisco v. 3, p. 1579-1589, 2005.

HU, X. *et al.* Paired chimpanzee hepatitis B virus (ChHBV) and mtDNA sequences suggest different ChHBV genetic variants are found in geographically distinct chimpanzee subspecies. **Virus Research**, Amsterdam, v.79, p. 103-108, 2001.

HUSSAIN, A.I. *et al.* Screening of simian foamy virus infection by using a combined antigen western blot assay: evidence for a wide distribution among old world primates and identification of four new divergent viruses. **Virology**, San Diego, v. 309, p. 248-257, 2003.

JAKAB, F. *et al.* Detection of Dobrava hantaviruses in *Apodemus agrarius* mice in the Transdanubian region of Hungary. **Virus Research**, Amsterdam, v. 128, p. 149-152, 2007.

JOHNSON, N. *et al.* Molecular epidemiology of canid rabies in Sudan: evidence for a common origin of rabies with Ethiopia. **Virus Research**, Amsterdam, v. 104, p. 201-205, 2004.

JÚNIOR, P.C. *et al.* Molecular epidemiology of rabies virus strains isolated from wild canids in Northeastern Brazil. **Virus Research**, Amsterdam, v. 120, p. 113-120, 2006.

JUNQUEIRA, L.C.U.; CARNEIRO, J. **Biologia celular e molecular**. Rio de Janeiro: Guanabara Koogan, 2005.

KINSELLA, E. *et al.* Sequence determination of the Crimean-Congo hemorrhagic fever virus L segment. **Virology**, San Diego, v. 321, p. 23-28, 2004.

KNOWLES, N.J.; *et al.* Molecular analysis of encephalomyocarditis viruses isolated from pigs and rodents in Italy. **Virus Research**, Amsterdam, v. 57, p. 53-62, 1998.

KUNO, G.; CHANG, G-J.J. Phylogeny of genus *Flavivirus*. **Journal of Virology**, London, v. 72, p. 73-83.

LECOMPTE, E. *et al.* Genetic identification of Kodoko Virus, a novel Arenavirus of the African pigmy mouse (*Mus Nannomys minutoides*) in West Africa. **Virology**, San Diego, v. 364, p. 178-183, 2007.

LESER, W. Elementos em epidemiologia geral. Rio de Janeiro: Atheneu, 1985.

LIEGEOIS, F. *et al.* Molecular characterization of a novel simian immunodeficiency virus lineage (SIVtal) from northern talapoin (*Miopithecus ougouensis*). **Virology**, San Diego, v. 349, p. 55-65, 2006.

MATIOLI, S.R. **Biologia molecular e evolução**. Ribeirão Preto: Holos, 2001.

MONCAYO, A.C. *et al.* Allpahuayo Virus: a newly recognized Arenavirus (*Arenaviridae*) from arboreal rice rats (*Oecomys bicolor* and *Oecomys paricola*) in Northeastern Peru. **Virology**, San Diego, v. 285, p. 277-286, 2001.

MORI, E. **Infecção experimental em cavalos pelo Herpesvírus Eqüino tipo 1: aspectos clínicos e detecção do agente pela reação em cadeia pela polimerase**. 2008. 158 f. Dissertação (Doutorado em Clínica Veterinária). Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo/SP.

O'BRIEN, S.J. *et al.* Plagues and adaptation: lessons from the felidae models for SARS and AIDS. **Biological Conservation**, v. 131, p. 255-267, 2006.

OLIVEIRA, R.N. **Vírus da raiva em morcegos insetívoros: implicações em epidemiologia molecular da diversidade dos genes codificadores da nucleoproteína e glicoproteína**. 2009. 81 f. Dissertação (Mestrado em Epidemiologia Experimental Aplicada às Zoonoses). Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo/SP.

PÁEZ, A.; VASCO-VILLA, A.; REY, G.; RUPPRECHT, C.E. Molecular epidemiology of rabies in Colombia 1994-2005 based on partial nucleoprotein gene sequences. **Virus Research**, Amsterdam, v. 130, p. 172-181, 2007.

PAGE, R.D.M.; HOLMES, E.C. **Molecular evolution: a Phylogenetic Approach**. Oxford: Blackwell, 1998.

PENA, H.F.J. *et al.* Isolation and molecular detection of *Neospora caninum* from naturally infected sheep from Brazil. **Veterinary Parasitology**, Amsterdam, v. 147, p. 61-66, 2007.

PISONI, G. *et al.* Compartmentalization of small ruminant lentiviruses between blood and colostrums in infected goats. **Virology**, San Diego, v. 369, p. 119-130, 2007.

PISONI, G. *et al.* Demonstration of coinfection with and recombination by caprine arthritis-encephalitis virus and maedi-visna virus in naturally infected goats. **Journal of Virology**, Washington, v. 81, n. 10. p. 4948-4955, 2007.

PISONI, G.; QUASSO, A.; MORONI, P. Phylogenetic analysis of small-ruminant lentivirus subtype B1 in mixed flocks: evidence for natural transmission from goats to sheep. **Virology**, San Diego, v. 339, p. 147-152, 2005.

PISONI, G. *et al.* Phylogenetic analysis of the *gag* region encoding the matrix protein of small ruminant lentiviruses: comparative analysis and molecular epidemiological applications. **Virus Research**, Amsterdam, v. 116, p. 159-167, 2006.

REAL, L.A.; BIEK, R. Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. **Journal of the Royal Society**, London, v. 4, p. 935-948, 2007

RILEY, LEE W. **Molecular epidemiology of infectious diseases: principles and practices.** Washington: ASM Press, 2004.

SCHNEIDER , H. **Métodos de Análise filogenética-** um guia prático. Bragança, Pará, 2003

SCHUMANN, T. *et al.* Evidences of interspecies transmission and reassortment among avian group A rotaviruses. **Virology**, San Diego, v. 386, p. 334-343, 2009.

SMITH, G.J.D. *et al.* Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. **Nature**, London, v. 459, p. 1122-1126, 2009.

SOUNIS, E. **Epidemiologia geral.** São Paulo: Atheneu, 1985

SU, C.;ZHANG, X.;DUBEY, J.P. Genotyping of *Toxoplasma gondii* by multilocus PCR-RFLP markers: A high resolution and simple method for identification of parasites. **International Journal for Parasitology**, Oxford, v.36, p. 841-848, 2006.

SWOFFORD, D.L. **PAUP Version 4**: phylogenetic analysis using parsimony (and other methods). Sunderland: Sinauer Associates, 2002.

TAJIMA, F., NEI, M. Estimation of evolutionary distance between nucleotide sequences. **Molecular Biology and Evolution**, Oxford, v. 1, pp. 269-285, 1984.

TAMURA, K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and GC-content biases. **Molecular Biology and Evolution**, Oxford, n. 9, p.678-687, 1992.

TAMURA, K., NEI, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. **Molecular Biology and Evolution**, Oxford, v. 10, n. 3, p. 512-526, 1993.

TURNER, P.C.; MOTTA, P.A.; **Biologia molecular**. Rio de Janeiro: Guanabara Koogan, 2004.

TROY, C.S. *et al.* Principles and methods of sequence analysis. **Nature**, London, v. 410, p. 1091-1110, 2001.

VANDAMME, A.M. **A practical approach to DNA and protein phylogeny**. Cambridge: Cambridge University Press, 2003.

VILLARREAL, L.Y.B. *et al.* Orchitis in rosters with reduced fertility associated with avian infectious bronchitis virus and avian metapneumovirus infectious. **Avian Diseases**, Kennett Square, v. 51, p. 900-904, 2007.

VILLARREAL, L.Y.B. *et al.* Molecular characterization of infectious bronchitis virus strains isolated from the enteric contents of brazilian laying hens and broilers. **Avian Diseases**, Kennett Square, v. 51, p. 974-978, 2007.

YANG, Z. Estimating the pattern of nucleotide substitution. **Journal of Molecular**, v. 39, p. 105-111, 1994.

WARNER, C.K.; SCHURR, T.G.; FEKADU, M. Molecular characterization of carrier rabies isolates. **Virus Research**, Amsterdam, v. 41, p. 133-140, 1996.

ZOU, Y. *et al.* Genetic analysis of hantaviruses carried by reed voles *Microtus fortis* in China. **Virus Research**, Amsterdam, v. 137, p. 122-128, 2008.