

Laboratório Nacional de Computação Científica
Programa de Pós Graduação em Modelagem Computacional

**Filodinâmica e Evolução Molecular dos Sorotipos de
Vírus da Dengue**

Por
Raquel Lopes Costa

PETRÓPOLIS, RJ - BRASIL

MAIO DE 2010

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

**FILODINÂMICA E EVOLUÇÃO MOLECULAR DOS SOROTIPOS
DE VÍRUS DA DENGUE**

Raquel Lopes Costa

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO LABORATÓRIO
NACIONAL DE COMPUTAÇÃO CIENTÍFICA COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM MODELAGEM COMPUTACIONAL

Aprovada por:

Prof. Carlos Guerra Schrago, D.Sc
(Presidente)

Prof. Marisa Fabiana Nicolás, Ph.D

Prof. Carolina Moreira Voloch, D.Sc.

Prof. Renato Santana de Aguiar, D.Sc.

PETRÓPOLIS, RJ - BRASIL
MAIO DE 2010

Lopes Costa, Raquel

XXXX Filodinâmica e evolução molecular dos sorotipos de vírus da dengue /
Raquel Lopes Costa. Petrópolis, RJ. : Laboratório Nacional de Computação
Científica, 2010.

xx, yy p. : il.; 29 cm

Orientador: Carlos Guerra Schrago

Dissertação (M.Sc.) – Laboratório Nacional de Computação Científica,
2010.

1. ASSUNTO. 2. palavra chave. 3. palavra chave. 4. palavra chave. I.
Guerra Schrago, Carlos. II. LNCC/MCT. III. Título.

CDD XXX.XXX

“Se você encontrar um caminho sem
obstáculos, ele provavelmente não leva a
lugar nenhum.”

(Autor: Frank Clark)

Resumo da Dissertação apresentada ao LNCC/MCT como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

FILODINÂMICA E EVOLUÇÃO MOLECULAR DOS SOROTIPOS DE VÍRUS DA DENGUE

Raquel Lopes Costa

Maio , 2010

Orientador: Carlos Guerra Schrago, D.Sc

Os vírus, especialmente os que apresentam RNA como material genético, possuem uma rápida taxa evolutiva, acumulando variação genética em um intervalo de tempo acessível a estudos epidemiológicos. Entre os vírus de RNA, o vírus da dengue (*Flavivirus*) é o agente da mais ampla doença viral transmitida por um inseto vetor do gênero *Aedes* que afeta países das regiões tropicais e subtropicais. Através de técnicas de inferência bayesianas por MCMC este estudo avaliou padrões de distribuição viral e evolução molecular dos quatro sorotipos de vírus da Dengue (DenvI-IV) em 603 amostras da região do envelope distribuídas em 71 países e isolados por um período de 64 anos (1944-2008). Os resultados mostraram semelhanças nas taxas médias de evolução dos sorotipos (em torno de 7.5 subst./sítio/ano) com uma considerável sobreposição de 95% HDP. A coalescência das amostras conjuntas sugere que a primeira divergência entre os sorotipos aconteceu há aproximadamente 1679 anos com processos de extinção, aparecimento e expansão de novas linhagens responsáveis pela diversidade intrasorotípica atualmente observada. As análises populacionais revelaram ocorrência de picos de crescimento relacionadas com explosivas epidemias coincidentes com introduções sorotípicas ocorridas principalmente nas Américas Central e do Sul. Padrões de distribuição geográfica mostraram forte estrutura espaço-temporal com agrupamentos principalmente no sudeste asiático e Américas.

Abstract of Dissertation presented to LNCC/MCT as a partial fulfillment of the requirements for the degree of Master of Sciences (M.Sc.)

TÍTULO EM INGLÊS DA TESE OU DISSERTAÇÃO

Raquel Lopes Costa

May, 2010

Advisor: Carlos Guerra Schrago, D.Sc

Sumário

1	Introdução	1
1.1	Organização da Dissertação	1
1.2	Vírus	2
1.3	<i>Flavivirus</i> (Flaviviridae)	3
1.4	Dengue	5
1.4.1	Sintomas e Fatores de Risco DHF/DSS	5
1.4.2	Transmissão & Replicação	6
1.4.3	Ciclos	7
1.4.4	Prevenção, Controle e Vacinas	8
1.5	Histórico da Virose	8
1.5.1	Dengue nas Américas	9
1.6	Diversidade Intrasorotípica e Taxas de Mutação	11
2	Objetivos	16
2.1	Objetivos Gerais	16
2.2	Objetivos Específicos	16
3	Materiais & Métodos	17
3.1	Conjunto de Dados	17
3.2	Evolução Molecular	19
3.2.1	Modelos Substituição de Nucleotídeos	19
3.2.2	A Inferência Bayesiana via MCMC	21

3.3	Padrões Filodinâmicos de Viroses Humanas de RNA	22
3.4	Análise da História Evolutiva	23
3.4.1	Genealogias	23
3.4.2	Diversidade Intrasorotípica	24
3.4.3	Taxas de Substituição de Nucleotídeos e TMRCA	25
3.4.4	Análise Populacional baseada em Variação de Sequência	26
3.4.5	Análise da Ação Diferencial de Seleção nos Códonos do Gene	26
4	Resultados	29
4.1	Estimativas de Taxa Substituição e TMRCA	29
4.2	Genalogia, Padrões de distrições, Diversidade intrasorotípica e Dinâmicas populacionais	31
4.2.1	Dengue I	31
4.2.2	Dengue II	33
4.2.3	Dengue III	35
4.2.4	Dengue IV	36
4.3	Indícios de Sítios sob Pressão Seletiva	38
5	Discussão	45
6	Conclusão & Perspectivas	52
	Referências Bibliográficas	53
	Apêndice	
A	Banco de Dados: DENV Database	62
A.1	Modelo Conceitual	62
A.2	Modelo Lógico	63
A.3	Modelo Físico	64
A.4	Dados e Estatísticas	65

A.5	Consultas	65
A.6	Portal Web	65
A.7	Ferramentas	66
B	Lista das Siglas e Países Utilizadas nas Análises de Acordo com a ISO 3166-1	68
C	Sequências do GenBank	70
D	Bases de dados Especializadas em Virus da dengue	75
E	Modelos Substituições de Nucleotídeos	76
F	Breve Descrição dos Programas Utilizados	79
F.1	BEAST	79
F.1.1	Beauti	79
F.1.2	TreeAnnotator	80
F.2	TRACER	80
F.3	PAML	80
F.3.1	CODEML	81
F.4	ClustalW	81
F.5	MEGA	81
G	Árvores Genealógicas Circulares	82

Lista de Figuras

Figura

1.1	Esquema do genoma de <i>Flavivirus</i> mostrando proteínas estruturais e não estruturais.	4
1.2	Distribuição geográfica mundial da Dengue em 2001. Fonte: Gubler (1998)	5
1.3	Ciclos de transmissão do vírus da dengue em insetos do gênero <i>Aedes</i> . Fonte: Gubler (2002)	8
1.4	Linha do tempo com os principais acontecimentos e registros epidêmicos relacionados à Dengue.	10
1.5	Distribuição de temporal e espacial <i>Aedes aegypti</i> durante as décadas de 30, 70 e 2000 nas Américas. Fonte: Gubler (2002)	12
3.1	Transições e Transversões ocorridas nos nucleotídeos (A,T,C,G	20
3.2	Ilustração de substituições multiplas ocorrendo em um mesmo sítio ou em multiplos sítios. Fonte: (Yang, 2006)	20
3.3	Padrões filogeográficos de víroses de RNA. Fonte: (Holmes, 2008) .	24
4.1	Taxas de substituição de nucleotídeos para as amostras analisadas.	30
4.2	Coeficiente de variação do relógio molecular relaxado.	30
4.3	Dinâmica populacional em Denv-I. A região sombreada representa 95% HDP e a linha sólida a estimativa média.	32
4.4	Dinâmica populacional em Denv-II. A região sombreada representa 95% HDP e a linha sólida a estimativa média.	34

4.5	Dinâmica populacional em Denv-III. A região sombreada representa 95% HDP e a linha sólida a estimativa média.	35
4.6	Dinâmica populacional em Denv-IV. A região sombreada representa 95% HDP e a linha sólida a estimativa média.	37
4.7	Árvore Filogenética para Denv-I.	40
4.8	Árvore Filogenética para Denv-II.	41
4.9	Árvore Filogenética para Denv-II.	42
4.10	Árvore Filogenética para Denv-IV.	43
4.11	Valores de ω para DenvI-IV respectivamente. Os pontos plotados em vermelhos indicam valores de ω com probabilidade posterior maior que 1 com intervalo de confiança.	44
A.1	Diagrama Entidade Relacionamento.	63
A.2	Esquema do Modelo Lógico.	64
A.3	Formulário de Consulta.	66
A.4	Aplicação Web.	66

Lista de Tabelas

Tabela

1.1	Estudos de substituição de nucleotídeos	15
3.1	Conjunto de Dados utilizado nas análises.	18
4.1	Estimativas de substituição de nucleotídeos	29
4.2	Resíduos sob pressão seletiva no modelo M0	38
4.3	Resíduos sob pressão seletiva no modelo M1	39
4.4	Resíduos sob pressão seletiva no modelo M2	39

Lista de Siglas e Abreviaturas

- BD: Banco de Dados;
- BEAST: *Bayesian Evolutionary Analysis by Sampling Trees*;
- BEB: *Bayes Empirical Bayes*;
- BLAST: *Basical Local Alignment Search Tool*;
- BSP: *Baysian Skyline Plot*;
- DENV: Vírus da Dengue (*Dengue Virus*);
- DF: Febre do Dengue (*Dengue Fever*);
- DHF: Febre Hemorrágica do Dengue (*Dengue Hemorrhagic Fever*);
- DNA: Ácido Desoxirribonucleico (*Deoxyribonucleic Acid*);
- DSS: Síndrome do choque por Dengue (*Dengue Shock Syndrome*);
- ENV: Envelope lipídico viral
- ER: Entidade-Relacionamento
- ESS: *Effective Sample Size*;
- GTR: *General Time Reversible*;
- HIV: Vírus da Imunodeficiência Humana (*Human Immunodeficiency Virus*);
- HKY: Hasegawa, Kishino e Yano;
- HPD: *Highest Posterior Density*;
- HTML: *HyperText Markup Language*;

- ISO: Organização Internacional de Normalização (*International Organization for Standardization*);
- JEV: Encefalite Japonesa (*Japanese Encephalitis Virus*);
- MCMC: Cadeia de Markov Monte Carlo (*Markov Chain Monte Carlo*);
- MRCA: Ancestral comum mais recente (*Most Recent Common Ancestor*);
- MySQL: SGBD com linguagem de consulta estruturada (*Structured Query Language*);
- NCBI: *National Center for Biotechnology Information*;
- OMS: Organização Mundial de Saúde (*World Health Organization, WHO*);
- PAHO: Organização Panamericana de Saúde (*Panamerican Health Organization*);
- PHP: *Hypertext Preprocessor*;
- RNA: Ácido Ribonucleico (*Ribonucleic Acid*);
- SGBD: Sistema Gerenciador de Banco de Dados;
- SQL: *Structured Query Language*;
- TMRCA: Tempo do mais Recente Ancestral Comum (*Time to Most Recent Common Ancestor*);
- W3C: *World Wide Web Consortium*;
- WNV: *West Nile Viruses*;
- YFV: Vírus da Febre Amarela (*Yellow Fever Virus*);

Capítulo 1

Introdução

A dengue é considerada uma das mais importantes doenças infecciosas emergentes e reemergentes no mundo (Klungthong et al., 2008). É causada por um vírus de RNA transmitido através do inseto vetor *Aedes* e embora existam quatro sorotipos distintos, Dengue I-IV, os sintomas provocados pela virose são os mesmos e variam desde um quadro assintomático e febre leve à hemorragias graves.

Registros históricos e controle epidemilógico da dengue são feitos desde o começo do século XX porém, apenas recentemente, informações acumuladas no genoma viral são usadas para reconstruir a história epidemiológica e prever seu curso. Tal fato deve-se a três principais avanços: aumento da quantidade e disponibilidade de sequências virais, melhoria do processamento computacional e desenvolvimento de sofisticadas técnicas estatísticas (Pybus e Rambaut, 2009). Tais estudos podem auxiliar na compreensão da história epidemiológica e também projetar áreas emergentes, possíveis epidemias, locais que são fontes e reservatórios virais além de padrões de dispersão e distribuição (Holmes, 2008).

1.1 Organização da Dissertação

O estudo proposto avaliou comparativamente todas as amostras dos quatro sorotipos de dengue encontradas para a região do envelope viral (ENV) disponíveis no GenBank até junho de 2009. A comparação foi feita com enfoque na evolução molecular e padrões de distribuição, através da reconstrução

genealógica, inferências de taxa de substituição de nucleotídeos (sítio/ano), tempo de divergência (anos), dinâmica populacional e pressões seletivas. Para melhor compreensão dos assuntos aqui tratados, a dissertação está dividida em 6 capítulos resumidos a seguir.

O capítulo 1 foi dedicado a uma breve descrição de vários aspectos relacionados ao vírus da dengue (DENV) como ciclos, severidade da infecção, aspectos epidemiológicos e históricos. Em seguida o capítulo 2 apresenta os objetivos deste estudo, tanto gerais quanto os mais específicos. No capítulo 3 é descrita a metodologias utilizadas nas análises evolutivas das sequências de dengue, dentre elas, o conjunto de dados montado, modelos de substituição de nucleotídeos, padrões filogeográficos, métodos e parâmetros dos modelos evolutivos. O capítulo 4 apresenta os vários resultados obtidos descritos separadamente para cada sorotipo. Já no capítulo 5 os sorotipos são discutidos envolvendo a observação de padrões gerais e padrões mais específicos. Finalmente no capítulo 6 é apresentada conclusões e perspectivas pertinentes ao estudo. Os diversos apêndices que compõem este trabalho são materiais de apoio ao texto inseridos de forma a auxiliar eventuais dúvidas que o leitor possa ter no decorrer da leitura.

1.2 Vírus

Os vírus são pequenos agentes infecciosos capazes de reproduzirem somente no interior de células de outros organismos. São encontrados em quase todos os ecossistemas e, desde a descoberta do vírus do tabaco por Beijerinck em 1898, cerca de 5.000 vírus foram descritos em detalhes até hoje (Dimmock et al., 2007) e, de acordo com Breitbart e Rohwer (2005) estima-se que haja 10^{31} vírus na Terra.

A transmissão de vírus é bastante variável e pode ocorrer de plantas para plantas (através de insetos como afídios), de animais para animais (através de mosquitos vetores), do contato direto entre pessoas contaminadas, como o influenza (por tosse ou espirro), rota oral-fecal como norovírus e rotavírus que causam gastroenterites ou pelo contato sexual como HIV e HPV.

Quanto a sua estrutura, os vírus são constituídos basicamente de um capsídeo e uma unidade reprodutora, o DNA ou RNA. Entram na célula por endocitose mediada por receptores celulares formando uma invaginação com formação de endossomas ou entram por fusão direta com a membrana do hospedeiro, sendo este último um mecanismo utilizado por alguns vírus com envelope (Aroldo, 2000).

Atualmente a classificação de vírus está baseada no sistema de Baltimore (Aroldo, 2000) agrupando os vírus de acordo com a síntese de RNA mensageiro e modo de replicação (Flint et al., 2004). Sendo assim, ocorrem sete classes de vírus: dsDNA (vírus DNA fita dupla), ssDNA (vírus DNA fita simples), dsRNA (vírus RNA fita dupla), (+) ssRNA (vírus RNA fita simples positivo), (-) ssRNA (vírus RNA fita simples negativo), ssRNA-RT (vírus RNA fita simples com DNA intermediário), dsDNA-RT (vírus DNA fita simples com RNA intermediário).

De modo geral, as epidemias e doenças humanas estão mais relacionadas a vírus de RNA do que os vírus de DNA (Holmes, 2004, 2008). Além disso, foi observado que taxas substituição de nucleotídeos ocorrem, em média, seis ordens de magnitude maiores que eucariotos e vírus de DNA (Jenkins et al., 2002). Erros provenientes da RNA polimerase (ausência do mecanismo de correção após duplicação gênica, *proofreading*) (Steinhauer, 1992), curto tempo de geração, tamanho do genoma e imenso tamanho populacional constituem os fatores responsáveis pelas altas taxas observadas.

Desta forma, estudos espaciais e temporais de epidemias humanas provocadas por vírus de RNA são possíveis dentro de uma escala de tempo ecológico não possível para os demais organismos (Grenfell et al., 2004).

1.3 *Flavivirus* (Flaviviridae)

Dentre os vírus de fita simples (ssRNA +), os flavivírus têm considerável importância epidemiológica pois contém cerca de 70 espécies onde mais de 50% delas estão associadas com doenças humanas tais como o vírus da dengue (*Dengue virus*, DENV), febre amarela (Yellow Fever virus, YFV), *West Nile viruses*,

(WNV), *Tick-Borne Encephalitis viruses*, (TBEV) e o vírus da encefalite japonesa (*Japanese Encephalitis virus*, JEV) (Weaver e Vasilakis, 2009). O tamanho dos flavivírus é relativamente pequeno variando de 40 a 60 nanômetros de diâmetro, com geometria esférica e envelopados (Gubler, 1998). Seu genoma caracteriza-se pela polaridade positiva e tamanho aproximado de 11kb (Figura 1.1) codificando proteínas estruturais e não estruturais localizados entre as terminações 5' (100 nucleotídeos metilados) e terminação 3' (400-800 nucleotídeos) que não apresenta uma cauda poliA. Estas terminações incluem regiões de seqüências não traduzidas que provavelmente interagem com fatores celulares envolvidos na replicação viral e tradução (Weaver e Vasilakis, 2009).

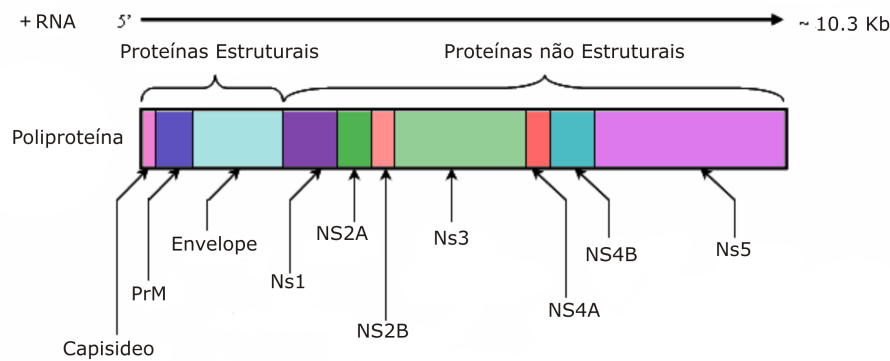


Figura 1.1: Esquema do genoma de *Flavivirus* mostrando proteínas estruturais e não estruturais.

Capsídeo (C), Membrana (M) e Envelope lipídico (ENV) constituem as proteínas estruturais, sendo esta última responsável por atividades biológicas do ciclo viral, incluindo montagem, interação com receptores celulares e fusão de membrana, além de ser o principal alvo de anticorpos neutralizantes e possuir atividade hemaglutinante Lanciotti et al. (1999); Petersen e Roehrig (2001) (Miagostovich, 1999). Dentre as proteínas não estruturais, NS1, NS3 e NS5 são de maior peso molecular e mais conservadas entre os flavivírus. Já as proteínas NS2A, NS2B, NS4A e NS4B são pouco conservadas, mas possuem domínios hidrofóbicos similares entre o gênero (Chambers et al., 1990).

1.4 Dengue

Incluído nos Flavivírus, o vírus da dengue é considerada a arbovirose (virose transmitidas por um artrópode hematófago a um vertebrado) mais difundida e de maior relevância médica (WHO, 2005; Weaver e Vasilakis, 2009). Estima-se que anualmente ocorram de 50 a 100 milhões de infecções em mais de 100 países das regiões tropicais e subtropicais (Figura 1.2) com aproximadamente 500.000 casos hospitalização e cerca de 2,5% de óbitos (WHO, 2005).

De acordo com (fulano) a tendência crescente dos casos está relacionada em grande parte a falta de vacinas, o aumento da urbanização, condições sócio-econômicas e mobilidade da população.

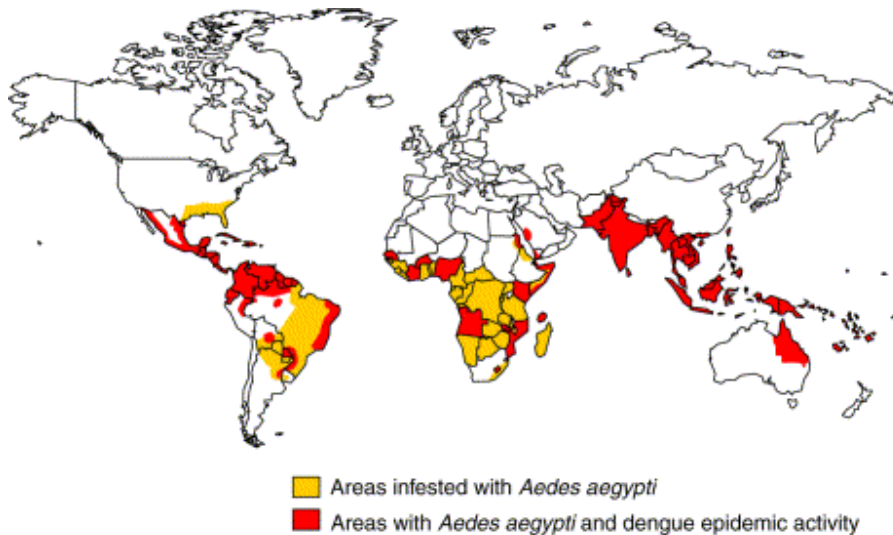


Figura 1.2: Distribuição geográfica mundial da Dengue em 2001. Fonte: Gubler (1998)

1.4.1 Sintomas e Fatores de Risco DHF/DSS

Infecções humanas sintomáticas podem resultar em uma febre mediana associada com dor de cabeça e dor retro orbital, mialgia, erupção cutânea, sintomas conhecidos como febre do dengue (*Dengue Fever*, DF) para uma forma mais severa da doença, associada com extravasamento do plasma, trombocitopenia

e manifestações hemorrágicas conhecida como dengue hemorrágica (*Dengue Hemorrhagic Fever*, DHF) ou ainda a síndrome de choque por dengue (*Dengue Shock Syndrome*, DSS) na qual ocorre extravasamento do fluido dentro do espaço intersticial resultando em choque hipovolêmico na qual sem um tratamento apropriado pode resultar em morte (Neto et al., 2005).

O mecanismo resultante de dengue hemorrágica e choque por dengue é ainda pouco compreendido (Kyle e Harris, 2008), mas sabe-se que há um conjunto de fatores envolvidos com este quadro agravante relacionados principalmente com fatores virais como sequências e existência de linhagens mais virulentas dentro dos sorotipos (RicoHesse et al., 1997; Guzman et al., 2007), fatores ambientais como densidade vetorial e populacional (Wilder-Smith e Gubler, 2008) e fatores do hospedeiro como infecções secundárias (Guzman et al., 2007), tempo entre infecções, sexo, idade, etnia e antecedentes genéticos do hospedeiro (Aroldo, 2000).

No aspecto dos fatores do hospedeiro, as infecções secundárias constituem a teoria conhecida como facilitação de infecção mediada por anticorpos (*Antibody Dependent Enhancement*, ADE), ou seja, os anticorpos resultantes da primeira infecção em concentrações subneutralizantes mediam a absorção do vírus da segunda infecção para o interior celular de monócito/macrófago via receptores Fc levando a um aumento da replicação viral e ativação imune acompanhado por uma liberação de citocina (Halstead et al., 2007). Em países onde há cocirculação de sorotipos (hiperendemicidade) a incidência de DHF/DSS é maior, corroborando com a teoria ADE.

1.4.2 Transmissão & Replicação

A passagem do vírus para o hospedeiro humano ocorre através da picada da fêmea do mosquito vetor que ao se alimentar de sangue infectado fornece um mecanismo de transmissão salivar. O vírus entra nas células hospedeiras por endocitose, mediado pela proteína viral do ENV com os receptores da membrana plasmática de células fagocitárias como macrófagos e monócitos que

são os principais sítios de replicação viral (Miagostovich, 1999).

É no citoplasma da célula ocorre a replicação viral. A cadeia complementar negativa de RNA é sintetizada por proteínas não estruturais sendo depois usada como molde para a síntese da progênie. A montagem ocorre durante o *budding* e ocorre preferencialmente em vacúolos citoplasmáticos e não na superfície celular como ocorre com outras arboviroses. O genoma viral é então traduzido como uma única poliproteína sendo depois clivada em proteínas maduras (C,M,ENV). A liberação dos vírus ocorre com a lise celular proliferando para tecidos musculares como as células musculares estriadas, lisas e fibroblastos, bem como linfonodos locais (Figueiredo, 1999).

1.4.3 Ciclos

São conhecidos dois ciclos de transmissão distintos: o endêmico/epidêmico e o enzoótico (Figura 1.3). O ciclo primeiro circula entre humanos servindo de reservatório e amplificação no hospedeiro. Neste ciclo, o *Ae. aegypti* é o principal transmissor urbano e está associado a epidemias explosivas que, em geral, têm início durante as estações chuvosas, quando o mosquito vetor existe em abundância (Weaver e Vasilakis, 2009). Já a espécie *Ae. albopictus* e outras espécies de *Aedes* veiculam principalmente a transmissão no meio rural.

O segundo ciclo de transmissão conhecido é denominado de florestal enzoótico ou ciclo silvestre, circulando entre primatas não humanos (servindo como reservatórios) e algumas espécies do gênero *Aedes* como subgêneros *Finlaya*, *Diceromyia* e *Stegomyia* (Gubler, 1998). Este ciclo ocorre principalmente em florestas na África Ocidental (Senegal, Nigéria, Burquina Faso e Guinéa) com prevalência de DENV-II e espécies de *Ae. aegypti formosus*, *Ae. luteocephalus*, *furcifer*, *taylori* e *vittalus*. Já nas florestas da região da Malásia (Wang et al., 2000) ocorre prevalência de Denv-I e Denv-IV. Embora o ciclo silvestre seja filogeneticamente distinto do ciclo endêmico/epidêmico, alguns registros evidenciam infecções humanas podem ocasionalmente ser provocadas por ciclos

silvestres (Kyle e Harris, 2008; Rudnick, 1966).

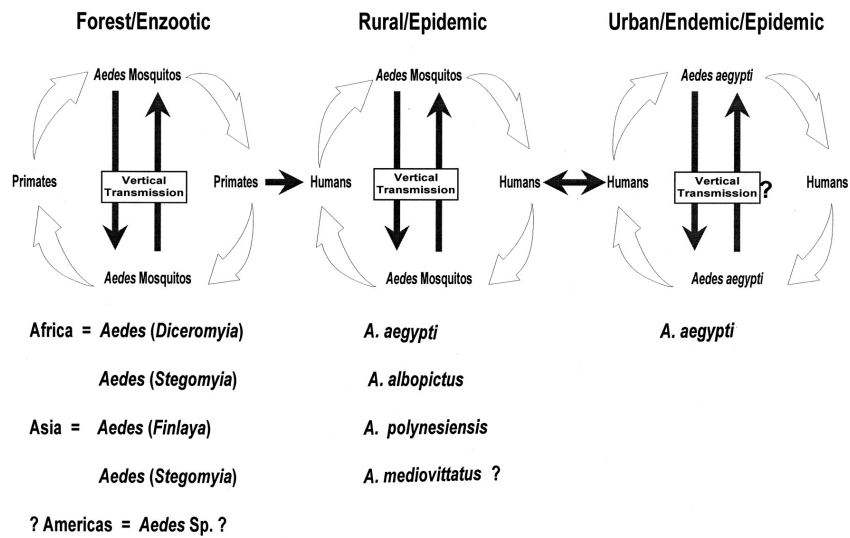


Figura 1.3: Ciclos de transmissão do vírus da dengue em insetos do gênero *Aedes*.
Fonte: Gubler (2002)

1.4.4 Prevenção, Controle e Vacinas

Atualmente nenhuma droga antiviral está disponível para o tratamento da dengue. Algumas vacinas estão em desenvolvimento, mas nada disponível para uso geral e um desafio particular com a vacinação DEN é evitar aumento da imunidade, o que poderia ocorrer se a vacina não protege contra todos os quatro sorotipos (Guy e Almond, 2008).

Desta forma, medidas de controle e prevenção da dengue são conduzidas, em grande parte, na limitação da distribuição do agente vetor (Kyle e Harris, 2008) com por exemplo, a aplicação de pesticidas em alguns focos de incidência epidêmica e campanhas de conscientização da população para eliminar fontes de água parada. Uma medida de controle recentemente apresentada por Bian et al. (2010) envolve o uso de bactérias *Wolbachia* que impedem o desenvolvimento do vírus no principal mosquito vetor, o *Aedes aegypti*.

1.5 Histórico da Virose

Os primeiros relatos de sintomas semelhantes ao do vírus da dengue aparecem na China durante as dinastias de Chin (265-420 DC), Tang (610 DC) e Sung do Norte (992 DC) (Gubler, 1998). Tais relatos estavam associados com a posição da água, insetos voadores e ao quadro clínico de febre e mialgia (Figura 1.4).

Durante os séculos XVII-XVIII foram registrados casos nas Ilhas francesas (1635), Panamá (1699), Jacarta (1799-1788), Cadiz e Sevilha na Espanha (1779), Cairo e Indonésia (1780). Até por volta de 1940 a epidemia foi marcada por baixa frequência e alta intensidade, e nenhuma comprovação evidenciara a enfermidade como sendo provocadas pelo vírus da dengue (Gubler, 1998).

A II Guerra Mundial trouxe devastação de enormes áreas florestais principalmente na Ásia e isto pode ter acarretado o aumento de surtos epidêmicos e hiperendemicidade nesta região Gubler (1998); Weaver e Vasilakis (2009). Neste mesmo período foram identificados e estabelecidos os sorotipos I e II provenientes da Tailândia e Filipinas (Kyle e Harris, 2008). Em 1953 durante epidemia ocorrida em Manila (Filipinas) aparece o primeiro registro de DHF (Hammon et al., 1960; Kyle e Harris, 2008) e identificação dos sorotipos III e IV. Depois de um período de 20 anos, a DHF se espalhou por todo sudeste asiático (Gubler, 1998).

Durante a década de 70 o vírus da dengue foi reintroduzido em vários países da Oceania (como ilhas Cook, Tahiti, Vanuatu, Palau e Nova Caledônia) após 25 de ausência. Em 1994 o vírus da dengue foi considerado a principal arbovirose sendo comparável a malária (Aroldo, 2000).

Devido a fiscalização limitada ou inexistente no continente africano, pouco se sabe sobre a história epidemiológica do Denv na região (Aroldo, 2000). O primeiro registro documentado ocorreu em 1927 na cidade de Durban (África do Sul). Os registros existentes estão concentrados principalmente da região oeste (Seicheles, Nigéria) e poucos registros constam a década de 80. Na década de 90 as epidemias são mais comuns no leste da África e meio leste com grande surto em Djibuti (1991) e em Jeddah, Arábia Saudita (1994).

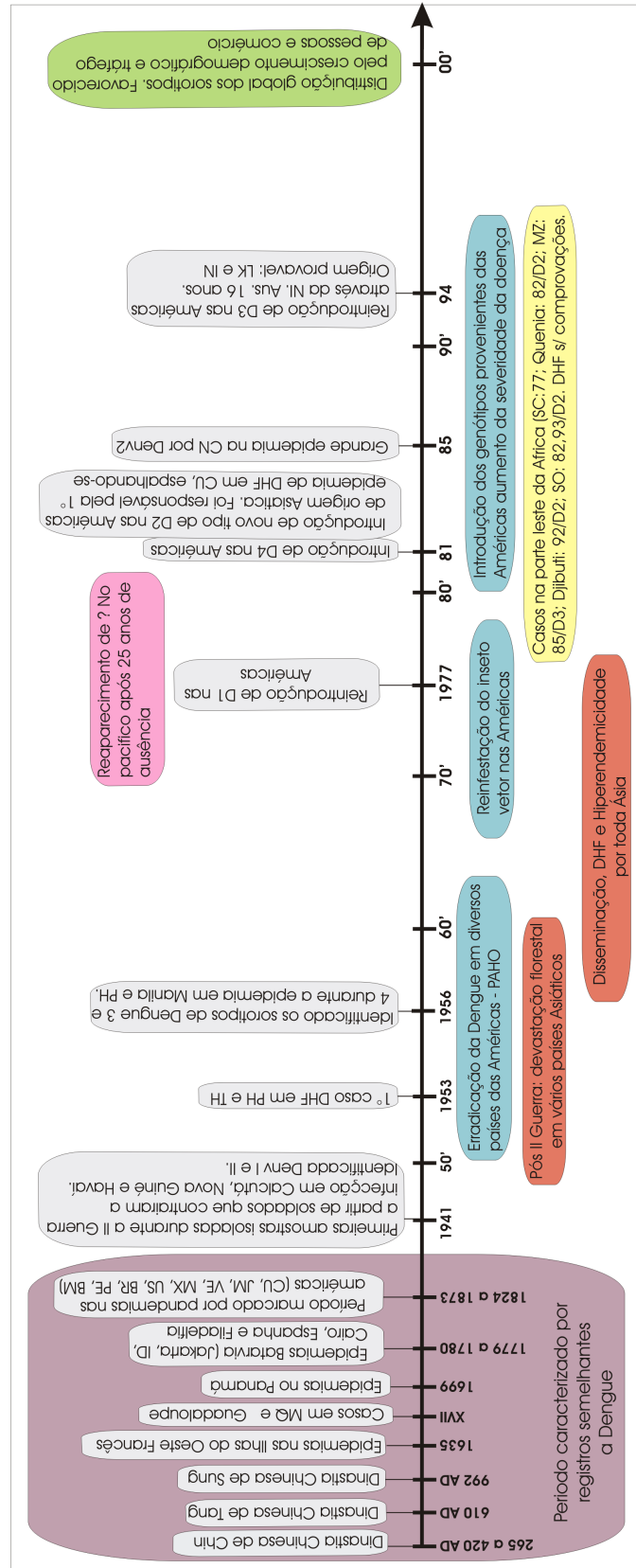


Figura 1.4: Linha do tempo com os principais acontecimentos e registros epidêmicos relacionados à Dengue.

1.5.1 Dengue nas Américas

Nas Américas os sintomas iniciais apareceram no século XVII em Martinica (Martinique) e Guadalupe (América Central). Em 1699 casos no Panamá e nos séculos XVIII e XIX na Filadélfia, US (1779), epidemias no Peru (1818), pandemias no golfo do Caribe, México, Colômbia, Rio de Janeiro, BR (1845), Bahamas (1882) e sul dos EUA (1850).

Durante as décadas de 50 e 60 a Organização Panamericana de Saúde (*Pan American Health Organization*, PAHO) promoveu uma campanha de erradicação de insetos transmissores de moléstias como febre amarela, malária e dengue. O êxito da campanha durou até a sua descontinuidade na década de 70, ocasionando a reinfestação do vetor juntamente com o vírus que ampliou sua distribuição geográfica (Figura 1.5) (Gubler, 1998).

Assim, novas linhagens de Denv-I foram introduzidas nas Américas na região do Caribe em 1977. Em 1981 ocorre reinfestação de Denv-II (Cuba) e Denv-IV (Porto Rico). Neste mesmo período ocorreu uma grande epidemia, considerada até então a maior epidemia de dengue nas Américas, localizada em Cuba com aparecimento da primeira DHF neste continente. Após alguns anos esta linhagem espalhou-se por vários países principalmente Venezuela, Colômbia, Brasil, Guiana Francesa, Suriname e Porto Rico. Em 1994 é reintroduzido o sorotipo Denv-III nas Américas após 16 anos de ausência tendo como portas de entrada Nicarágua e Panamá com cepas de origem do Sri Lanka e Índia (Aquino et al., 2009). Em 1998 e 1999 o Denv-III é introduzido nos países caribenhos como Porto Rico, Barbados, Jamaica e Martinique e em 2000 aparece na América do Sul.

Durante a década de 70 até os dias atuais muitos países das Américas passaram de um quadro não endêmico ou hipoendêmico (baixa co-ocorrência de sorotipos) para hiperendêmicos e com grandes epidemias de DHF, tanto quanto havia no sudeste asiático 25 anos antes (Gubler, 1998).

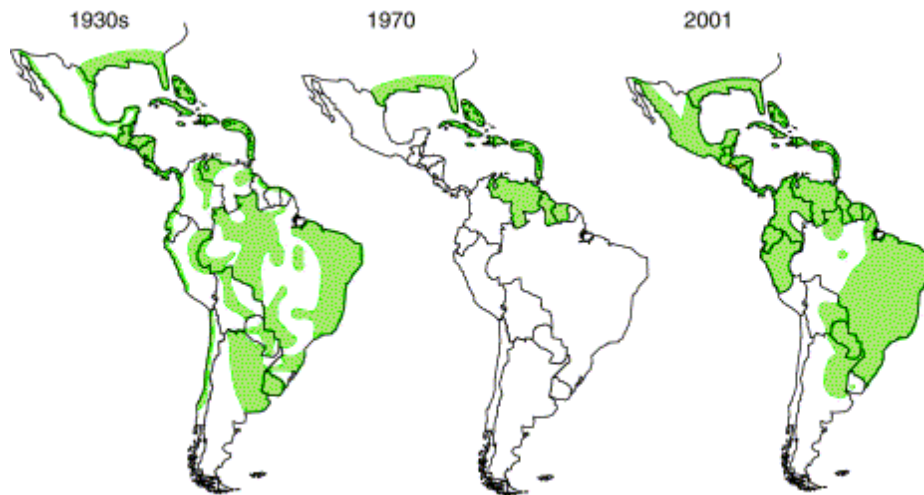


Figura 1.5: Distribuição de temporal e espacial *Aedes aegypti* durante as décadas de 30, 70 e 2000 nas Américas. Fonte: Gubler (2002)

1.6 Diversidade Intrasorotípica e Taxas de Mutação

Estudos prévios demonstram múltiplas linhagens (genótipos) entre os sorotipos de Dengue com algumas linhagens isoladas geograficamente e outras mais cosmopolitas.

Goncalvez et al. (2002) e Rico-Hesse (1990) reconheceram a existência cinco genótipos distintos em Denv-I utilizando a região de junção E/NS1 e posteriormente confirmada com o gene ENV. O genótipo I correspondeu a amostras do sudeste asiático, China e oeste da África. O tipo II agrupa amostras tailandesas da década de 50 e 60. O tipo III abrange sequências da Malásia, o tipo IV ilhas oeste do pacífico e Austrália e o tipo V compreende as Américas, oeste da África e um número limitado na Ásia.

A-Nuegoonpipat et al. (2004) e Kukreti et al. (2009) utilizando a mesma região gênica estabeleceram três genótipos distintos para Denv-I. O tipo I correspondente a linhagens de Mianmar, Tailândia, China, Malásia, Cingapura, Vietname Cambodja e algumas Africanas como Djibuti. O tipo II é classificado com sequências do pacífico: Austrália, Nova Caledônia, Salomão, Taiti, Timor Leste, Samoa ocidental e uma minoria na Ásia como Malásia, Indonésia, Filipinas

e Cingapura. O genótipo III é considerado o mais cosmopolita com amostras Americanas, partes da África e algumas de Mianmar. Amostras não compatíveis foram provenientes do Havai 1944, Japão 1943 e Tailândia 1963.

Vários trabalhos em Denv-II (Vasilakis et al., 2007; Twiddy et al., 2003a; Wang et al., 2000; RicoHesse et al., 1997; Ronquist e Huelsenbeck, 2003; Lewis et al., 1993; Twiddy et al., 2002) (Shurtleff 2001) estabeleceram cinco linhagens distintas. O tipo I foi representado por amostras provenientes da Guiné e Sri Lanka. O tipo II amostras das Filipinas e Taiwan. Tipo III compreendeu Tailândia, Malásia, Jamaica e Brasil. Tipo IV amostras asiáticas como Indonésia, Sri Lanka, e africanas como Somália, Seicheles, Burquina Faso. O tipo V teve sequências americanas de Trindade, Porto Rico, Tonga e asiáticas da Indonésia.

A filogenia de Denv-III foi apresentada nos trabalhos de Wittke et al. (2002); Lanciotti et al. (1994); Araujo et al. (2009) classificando o sorotipo III em cinco genótipos. O tipo I circula no sudeste asiático como Filipinas e em ilhas do sul do Pacífico. O tipo II amostras são provenientes do sudeste asiático. O tipo III foi cosmopolita com representantes asiáticos, africanos e americanos. O genótipo III circulante nas Américas teve origem provável do Sri Lanka de acordo com (Messer et al., 2003). O tipo IV constou América central como Porto Rico e Taiti.

O sorotipo de Denv-IV é considerado o mais conservado entre os demais. Estudos prévios separaram o Denv-IV em dois genótipos: o tipo I das Filipinas, Tailândia e Sri Lanka e o tipo II: Indonésia, Taiti, Caribe central e América do Sul (Lanciotti et al., 1994). Atualmente a filogenia de Denv-IV está separada em quatro sorótipos. O tipo I foi representado linhagens da Tailândia, Filipinas, Sri Lanka e Japão. O tipo II linhagens da Indonésia, Malásia, Taiti, caribe e Américas, o tipo III recentes linhagens da Tailândia e o tipo IV linhagens silvestres da Malásia (Klungthong et al., 2008).

Com relação as taxas de evolutivas estimadas para o vírus da dengue, alguns estudos foram conduzidos por diferentes autores, empregados métodos diferentes que variam desde medidas simples de regressão logística como propostas

por Lanciotti et al. (1997) e Zanotto et al. (1996) a métodos mais sofisticados empregando máxima verossimilhança ou métodos bayesianos (Tabela 1.1). O número amostral é variável devido ao tempo do estudo e/ou a regiões geográficas específicas. Os valores mais baixos encontrados de substituições de nucleotídeos foram para Denv-I ($4.55 * 10^{-4}$ substituições/sítio/ano) mostrado nos trabalhos de Twiddy et al. (2003a) que também apresentou maiores valores para Denv-III ($9.01 * 10^{-4}$ subst./sítio/ano).

Embora o vírus da dengue seja estudado sob o enfoque evolutivo, ainda não foi realizada nenhuma análise comparativa utilizando técnicas de inferência para parâmetros evolutivos, dinâmicas epidemiológicas e avaliação de dinâmicas espaço-temporais.

Tabela 1.1: Estudos de substituição de nucleotídeos

Sorotipo	Região	Qtd	sub/sítio/ano x10 ⁻⁴	Método	m.substi	TMRCa	autor/ano
D1	ENV	24	4.9(1.9-7.9)	ML	HKY85	-	Jenkins et al. (2002)
D2	ENV	24	6.2(4.9-7.4)	ML	HKY86	-	Jenkins et al. (2002)
D3	ENV	24	7.2(5.3-9.1)	ML	HKY87	-	Jenkins et al. (2002)
D4	ENV	24	7.7(5.9-10)	ML	HKY88	-	Jenkins et al. (2002)
D1	ENV	5	4.42	ML	HKY89	170	Zanotto et al. (1996)
D2	ENV	5	2.52	-	-	215	Zanotto et al. (1996)
D3	ENV	3	6.64	-	-	178	Zanotto et al. (1996)
D4	ENV	2	8.54	-	-	119	Zanotto et al. (1996)
D4	ENV	-	8.3	parcimonia	-	-	Lanciotti et al. (1997)
D1	ENV	9	4.55(2.54-6.54)	ML	GTR+G	126	Twiddy et al. (2003a)
D2	ENV	54	6.07(5.08-7.09)	ML	GTR+G	320.9	Twiddy et al. (2003a)
D3	ENV	21	9.01(7.27-10.84)	ML	GTR+G	100.89	Twiddy et al. (2003a)
D4	ENV	20	6.62(4.29-8.07)	ML	GTR+G	195.32	Twiddy et al. (2003a)
D3	ENV	200	8.9(7.9-10)	bayesiana	GTR+G+I	1891	Araujo et al. (2009)
D2 tipoIII	-	103	8.0(6.6-9.5)	bayesianos	GTR+G+I	-	Carrington et al. (2005)
D4 tipo II	-	99	8.3(6.8-10)	bayesianos	GTR+G+I	-	Carrington et al. (2005)
D4	ENV	47	10.72(8.41-13.11)	-	GTR+G+I	-	Klungthong et al. (2004)

Capítulo 2

Objetivos

2.1 Objetivos Gerais

Sob o enfoque da evolução molecular e filodinâmica este estudo teve como objetivo geral analisar **comparativamente** os quatro sorotipos de vírus da dengue.

2.2 Objetivos Específicos

Estimar as taxas de substituições de nucleotídeos e tempo de coalescência dos sorotipos isolados e em conjunto;

Caracterizar a diversidade intrasorotípica atual verificando linhagens emergentes bem como linhagens extintas;

Observar dinâmicas populacionais e relacionar com principais surtos epidêmicos da virose;

Verificar a heterogeneidade de ação da seleção ao longo dos códonos do gene;

Reconstruir genealogias e associar estruturas espaço-temporais através de filodinâmica;

Construir uma base de dados para seleção de sequências virais de DENV;

Observar a relação existente entre linhagens e potenciais emergentes de DHF/DSS;

Capítulo 3

Materiais & Métodos

3.1 Conjunto de Dados

Em todas as análises foi utilizado a região completa do envelope lipídico que contém cerca de 1485 pares de bases (pb). As sequências foram extraídas do banco de dados relacional Denv-Database construído a partir de sequências disponibilizadas no GenBank até junho de 2009.

As principais informações utilizadas em estudos de evolução molecular viral são, além do gene específico, o ano de coleta, localidade de origem e, em alguns casos, sorotipo determinado. No estudo de evolução molecular do vírus da dengue, todos estes aspectos são relevantes e infelizmente, muitos dados contendo estas informações encontravam-se em campos incorretos ou eram redundantes no GenBank justificando a montagem de uma base de dados para este estudo. Antes de utilizar o GenBank foi buscado bases de dados especializadas para o vírus da dengue disponíveis na web (Apêndice D) porém as informações contidas nestas bases estavam ou desatualizadas ou com distribuição geográfica restrita.

Foi escolhido o modelo de dados relacional na modelagem do Denv-Database devido as vantagens de simplificação da tarefa de programação, segurança e consistência dos dados (Aroldo, 1990). O banco Denv-Database resultou em 4 tabelas relacionais principais (sequências, genes, país, referência) com informações do número de acesso no Genbank, tamanho da sequência, ano de isolamento, país, gene, dados bibliográficos. Os detalhes da modelagem do banco podem ser visto

no apêndice A.

Após populado o banco de dados, o conjunto de dados escolhido Neste estudo a extração das sequências da base de dados obedeceu aos seguintes critérios:

- (a) Sequências com dados de ano, país e sorotipo definidos;
- (b) Sequências completas do envelope lipídico, ENV;
- (c) Sequências conhecidas como recombinantes foram excluídas das análises;
- (d) Com exceção das amostras consideradas mais antigas (antes de 1980), não foram repetidos o país e o ano da sequência;

Deste modo obteve-se um total de 603 sequências referente a 71 países e isoladas por um período de 64 anos (Tabela 3.1). As sequências foram codificadas respectivamente, com sorotipo, ano de extração da amostra, país de origem (representado pelo código de duas letras estabelecido pela ISO 3166-1, Anexo B) e número de acesso no Genbank (Anexo C). Após esta etapa, os quatro sorotipos foram alinhados com uso do algoritmo ClustalW (Thompson et al., 1994) e feitas correções posteriores no programa MEGA v.4.1 (Kumar et al., 2008).

Tabela 3.1: Conjunto de Dados utilizado nas análises.

Sorotipo	num. sequências	num. países	Intervalo (anos)	Tm (pb)
Denv-I	143	39	1944 a 2008	1485
Denv-II	207	47	1944 a 2008	1485
Denv-III	163	37	1956 a 2008	1485
Denv-VI	90	28	1956 a 2008	1480

Para facilitar as análises dos padrões geográficos, os países do conjunto de dados foram organizados por continente estabelecendo quatro grandes grupos: Américas (Central, Sul e do Norte), África (incluindo Arábia Saudita), Ásia (Sudeste Asiático, Extremo Oeste e Meridional) e Oceania.

Os sorotipos foram analisados separadamente na maioria das estimativas. Entretanto para tempo de coalescência (TMRCA) e taxa substituição também foi realizada uma análise conjunta dos sorotipos.

3.2 Evolução Molecular

A evolução molecular está moldada por forças de seleção, deriva e mutações observadas através de moléculas como os nucleotídeos compostos por **A**denina, **G**uanina, **C**itosina, **T**imina. Combinações formadas pela sequência de três nucleotídeos (códon) resultam em um aminoácido.

Se analisarmos as possibilidades de combinações de códons vemos que é possível a formação de 64 combinações distintas, mas na natureza são formados apenas 20 aminoácidos através destas combinações. Tal fenômeno acontece porque ocorre redundância do código genético, ou seja, algumas combinações geram um mesmo aminoácido.

As mudanças de uma purina (adenina ou guanina) para outra purina ou de uma pirimidina (citosina ou timina) para outra pirimidina são as transições e as mudanças de uma purina para uma pirimidina e vice-versa se referem a transversões Figura 3.1.

Quando as alterações em regiões codificantes do genoma alteram o aminoácido formado são denominadas de sinônimas e substituições que não alteram o aminoácido codificado são não-sinônimas. Há ainda alterações que podem resultar na produção de um códon de parada (*stop codon*) conhecidas como mutações sem sentido (*no sense*) e na maioria das vezes, geram produtos não funcionais.

3.2.1 Modelos Substituição de Nucleotídeos

Quando se compara duas sequências homólogas, ou seja, mesma origem evolutiva, o número de diferenças observadas entre elas cresce à medida que aumenta o tempo de separação das mesmas. Em teoria, a taxa de substituições entre os sítios acumula-se linearmente com tempo na distancia genética. Entretanto substituições multiplas e paralelismo são frequentes o que torna necessário o uso de modelos de substituição específicos. Figura /refsubstMultiplas

O primeiro e mais simples modelo de substituição de nucleotídeos foi proposto

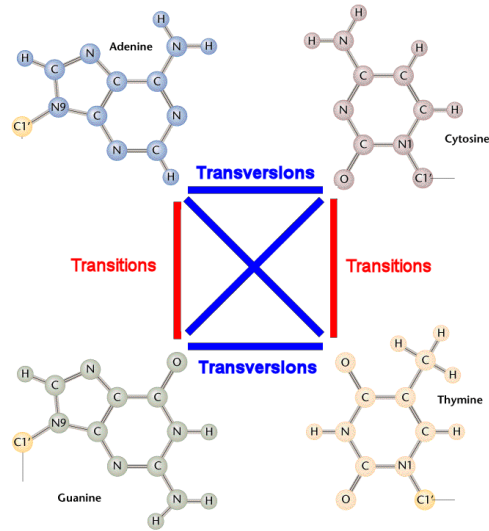


Figura 3.1: Transições e Transversões ocorridas nos nucleotídeos (A,T,C,G). Fonte: http://www.mun.ca/biology/scarr/Transitions_vs_Transversions.html.

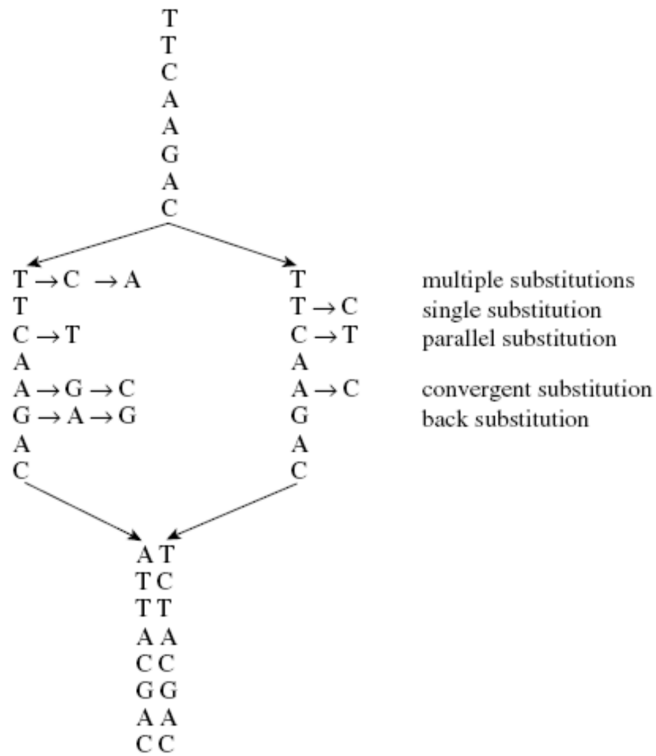


Figura 3.2: Ilustração de substituições múltiplas ocorrendo em um mesmo sítio ou em múltiplos sítios. Fonte: (Yang, 2006)

por Jukes e Cantor (1969) conhecido como JC69. Este modelo assume frequências iguais para as quatro bases ($A = C = G = T$) e igual probabilidade de substituição

para elas. A partir deste modelo, outros mais complexos incluíram a estimativa de um maior número de parâmetros como, por exemplo, o modelo GTR (Yang et al., 1994) na qual considera diferenças nas frequências das bases observadas ($A \neq T \neq C \neq G$). Maiores detalhes dos outros modelos de substituição de nucleotídeos encontram-se no Anexo F.

$$P_t = \begin{bmatrix} . & \pi_{Ca} & \pi_{Gb} & \pi_{Tc} \\ \pi_{Aa} & . & \pi_{Gd} & \pi_{Te} \\ \pi_{Ab} & \pi_{Cd} & . & \pi_{Tf} \\ \pi_{Ac} & \pi_{Ce} & \pi_{Tf} & . \end{bmatrix} f = [\pi_A \pi_C \pi_G \pi_T] \quad (3.1)$$

Geralmente os modelos propostos consideram a mesma taxa de substituição para os sítios. Entretanto, a primeira, segunda e terceira posição do códon variam e as análises estatísticas devem levar em consideração esta diferença (fulano). Por esta razão é incorporada neste modelo a distribuição gama derivada pelo mesmo método matemático que é usado na derivação das distâncias em sequências de aminoácidos de organismos distintos.

3.2.2 A Inferência Bayesiana via MCMC

A biologia molecular é influenciada por dados complexos. A maioria das propriedades genéticas dos indivíduos, populações e espécies são um produto de forças estocásticas e estudadas através do auxílio dos modelos probabilísticos Beaumont e Rannala (2004). Abordagens que consideram o uso de métodos bayesianos são recentemente utilizadas e diferem das abordagens clássicas chamadas de frequentistas. O atrativo da técnica bayesiana consiste em incorporar o conhecimento prévio de informações dentro de um modelo específico.

Assim, as técnicas computacionais MCMC de inferência bayesiana combinam informações dos parâmetros do modelo com informações dos dados (sintetizados pela função de verossimilhança) para formar a probabilidade posterior.

falta: importancia da simulacao estocastica via mcmc. Verossimilhança e

prior

3.3 Padrões Filodinâmicos de Viroses Humanas de RNA

Até pouco tempo atrás, a filogenia era usada na compreensão de padrões e processos que moldaram as populações ou diversificam espécies. Devido a quantidade de dados genéticos virais acumulados em grandes bases de dados biológicas de diversos países e em épocas diferentes, a filogenia passou a não olhar somente o passado, mas ser capaz de auxiliar controles epidemiológicos pois....Este estudo dinâmico com dados do passado que auxiliam medidas de controle futuro é usado como filodinâmica. A filodinâmica envolve análise da genealogia do gene no contexto espacial e infere processos históricos que formam a atual estrutura populacional e distribuição observada. Além da importância histórica destes estudos, eles ser usados como auxílio nas medidas de controle epidemiológico pois são capazes de fornecer informações de ocorrência de reservatórios, localidades emergentes e de difusões futuras para populações de hospedeiros (Holmes, 2008).

Algumas viroses de RNA em humanos já estudadas forneceram bases para estabelecer padrões gerais observados e indicados no trabalho de Holmes (2008). Além disto, estas viroses são altamente interconectáveis. Dentre elas: sem estrutura espacial, transmissão como onda, satélite central, dinâmica gravitacional e subdivisão populacional (Figura 3.3).

a) **Nenhuma estrutura espacial clara:** Ocorre um padrão complexo e difuso das amostras em diferentes regiões geográficas. Indicativo de fluxo viral frequente entre as localidades.

b) **Transmissão de onda:** O vírus move-se para fora de um ponto central de partida, produzindo, assim, uma relação relativamente simples entre a distância geográfica e genética. Isso é elegante exemplificado pelo vírus do sarampo (Grenfell et al., 2001).

c) **Fonte-dreno (ou satélite-centro).** Um número limitado de zonas geográficas exportam linhagens virais para outras regiões onde os vírus só podem

sobreviver no curto período (talvez por fatores sazonais).

d) **Dinâmica gravitacional.** Padrões de transmissão viral são impulsionados pelos centros de maior população, que atuam como atratores, talvez seguindo um padrão de fluxo de trabalho humano.

e) **Subdivisão espacial forte.** Diferentes localidades geográficas são caracterizadas pela presença de linhagens distintas filogeneticamente com pouca evidência de circulação viral entre eles. Apesar de subdivisão espacial forte ser esperada em vírus associados com infecções humanas de longos períodos, ela também pode ser estabelecida somente por um passado recente, dada a combinação de rápida evolução e os efeitos fundadores fortes, como visto em HIV.

Os padrões espaciais descritos são dependentes da escala espacial considerada, sendo possível que uma única virose passe por dois ou mais padrões filogeográficos. Além disso, novas linhagens são regularmente produzidas por mutação, mas a maioria é extinta de forma relativamente rápida e sem progresso para poucos períodos subsequentes de amostragem (Klungthong et al., 2008).

3.4 Análise da História Evolutiva

3.4.1 Genealogias

Na década de 80 estudos teóricos da variação genética de sequências moleculares através de processos estocásticos Griffiths, ano) Kingman (1982); Tavaré (1984); Tajima (1983); Hudson (1983) conhecidos como coalescência que resultaram em profundos avanços nos modelos clássicos de genética de populações (Rosenberg e Nordborg, 2002).

A teoria da coalescência descreve a distribuição de probabilidade dos tempos dos alelos encontram seu ancestral comum, ou seja, coalescem. Geração por geração a árvore genealógica é então construída até detectar coalescências de linhagens e finalmente o ancestral comum mais recente (*Time to the Most Recent Common Ancestor*, TMRCA) nas amostras. Neste processo, é possível inserir diferentes cenários demográficos e inferir taxas de crescimento populacional.

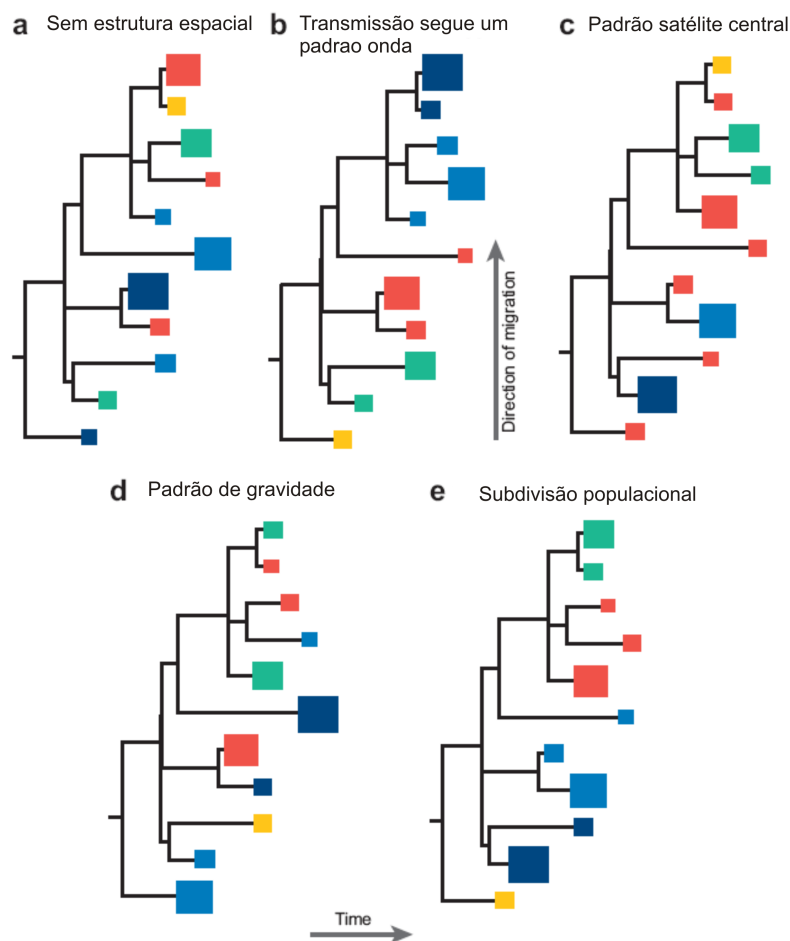


Figura 3.3: Padrões filogeográficos de vírozes de RNA. Fonte: (Holmes, 2008)

Neste estudo, a coalescência das amostras virais foi realizada através de métodos de reconstrução filogenética a partir do ano em que foram extraídas.

3.4.2 Diversidade Intrasorotípica

O termo genótipo usado em virologia e nesta dissertação, refere-se a uma classe similar, mas não idêntica de vírus isolados. O termo linhagem ou clado são usados como sinônimos dando o mesmo sentido descrito de indivíduos relacionados que compartilham um ancestral comum exclusivo (grupo monofilético) (Aroldo, 1990).

A alta diversidade observada em vírus de RNA permite agrupar as amostras em genótipos distintos possibilitando a inferência de padrões geográficos e

temporais. Ressalta-se que tal classificação é susceptível à mudança devido ao aumento da amostragem, localidades e revisões com sequências mais recentes.

Através da reconstrução genealógica das amostras e com base nos estudos de filodinâmica foi estabelecido genótipos para cada sorotipo de dengue e inferido padrões de distribuição observados.

3.4.3 Taxas de Substituição de Nucleotídeos e TMRCA

Uma das medidas centrais em evolução molecular é o estudo das taxas de substituição de nucleotídeos (Jenkins et al., 2002). Quando analisa uma árvore filogenética com tamanhos de ramos diferenciados infere-se relações de distâncias entre linhagens não sendo possível separar as taxas moleculares do tempo de divergência entre as amostras. O uso de técnicas como o relógio molecular relaxado assume que cada ramo tem uma taxa independente de evolução e com base em dados fornecidos como ano da amostra (utilizado para víruses) ou calibrações do registro fóssil é possível estimar taxas de evolução moleculares.

As estimativas das taxas evolutivas do vírus da dengue (substituição de nucleotídeo/sítio/ano, tempo de ancestral comum TMRCA em anos) foram feitas através de estimativas bayesianas com reconstrução genealógica. Foi utilizado o método Cadeia de Markov Monte Carlo (*Metropolis-Hasting Markov Chain Monte Carlo*, MCMC) disponível no pacote BEAST versão 1.5.3 (Drummond e Rambaut, 2007). O método MCMC é utilizado para construir a probabilidade posterior com estacionalidade após um período de *burn-in*. Em todos os parâmetros aplicou-se o relógio molecular relaxado (uncorrelated lognormal) descrito por Drummond et al. (2006) e o modelo de substituição de nucleotídeos *General Time Reversible*, GTR + Gama G que é consistentemente melhor relatado para evolução viral de acordo com (Zhou e Holmes, 2007). O número de gerações empregados na cadeia de Markov foi de 30.000.000 ciclos.

As estimativas de incertezas em torno dos parâmetros estimados ficaram em torno do intervalo 95% de probabilidade posterior (*Highest Posterior Density*,

HPD) visualizados no programa TRACER (v.1.5). Os softwares TreeAnnotator e FigTree permitiram a reconstrução da genealogia (com burn-in de 10%) e a visualização dos dados, respectivamente.

3.4.4 Análise Populacional baseada em Variação de Sequência

Os modelos demográficos tradicionais utilizam funções matemáticas para descrever mudanças no tamanho efetivo populacional ao longo do tempo. Dentre os mais simples e comuns estão o modelo de crescimento constante (um parâmetro), exponencial (dois parâmetros) e o logístico (população decresce depois de um tempo devido a sua capacidade de suporte, três parâmetros) (Drummond 2005). Modelos mais complexos utilizam os modelos mais simples incluindo outras variáveis a serem estimadas. O uso de um modelo demográfico incorreto influencia estimativas inválidas da história demográfica. Atualmente é possível traçar a dinâmica populacional no tempo utilizando amostras de sequências moleculares sem a dependência de um modelo paramétrico específico (Strimmer 2001).

Neste estudo utilizou-se o modelo *Bayesian Skyline Plot* (BSP) disponível no software TRACER. O BSP é um método que estima a dinâmica da população ancestral de uma amostra de sequências moleculares. O modelo BSP não requer parâmetro demográfico prévio Drummond et al. (2003, 2005) e permite encontrar novas assinaturas demográficas que não são facilmente descritas por modelos mais simples. O método leva em consideração tanto o erro inerente na reconstrução filogenética quanto o erro estocástico intrínseco do processo de coalescência produzindo estimativas mais corretas de incerteza estatística. Figura 3.4

3.4.5 Análise da Ação Diferencial de Seleção nos Códonos do Gene

Ao observar a frequência de um alelo ao longo do tempo pode-se inferir a ação e tipo de seleção que atua sobre ele. Em geral, substituições não sinônimas são deletérias e alteram o funcionamento da proteína gerada. Quando substituições não sinônimas não alteram a forma e função de uma proteína a força mais expressiva

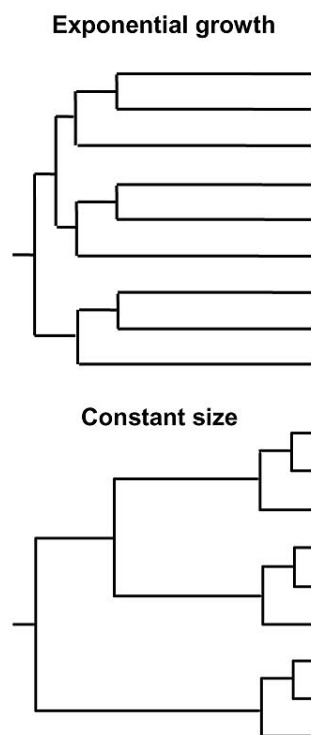


Figura 3.4

é de mutações neutras. Alterações que provocam uma maior frequência da proteína nova do que o esperado ao acaso confere o aparecimento de características adaptativas, um padrão conhecido como seleção positiva.

Em populações, a variação encontrada é maior do que a esperada caso fosse influenciada pela seleção natural e são fixadas por variações genéticas aleatórias (deriva gênica). Esta observação levou a proposição de teoria neutra proposta por Kimura (1968) na qual sugere que as chamadas mutações vantajosas possuem pouca contribuição para a variabilidade genética das populações por serem extremamente raras e se fixarem muito rapidamente pela seleção natural. Segundo Kimura, a variabilidade genética é uma fase transitória até a fixação do alelo por deriva gênica, ou seja, a maioria dos novos alelos inseridos na população são neutros.

Para determinar a natureza de pressões seletivas atuando no gene ENV, particularmente a presença de seleção positiva, foram realizadas análises comparativas de substituições sinônimas (dS) e não sinônimas (dN) através da

taxa $w = dN/dS$ na qual indica ocorrência de evolução neutra neutra ($w = 1$), purificadora ($w < 1$) ou positiva ($w > 1$).

Entretanto a taxa média sobre todos os sítios e em todas as linhagens quase nunca é maior que 1 uma vez que é improvável que a seleção positiva afete todos os sítios sob um tempo prolongado. Assim, para detectar seleção positiva a taxa ω , pode ser estimada por sítio ou através de linhagens utilizando modelos específicos como M0, M1 e M2 (Yang et al., 2000).

O modelo M0 (uma taxa) assume que a taxa w é igual em todas as sequências de códons. É o modelo mais simples e pode ser usado para checar se parâmetros estimados nos modelos mais complexos são consistentes. O modelo M1 (neutro) assume duas classes de sítios com $w_0 < 1$ estimado dos dados e $w_1 = 1$. O modelo M2 (seletivo) adiciona uma terceira classe de sítios a M1 com $w > 1$ estimado dos dados.

Todos os modelos de seleção envolvem os seguintes parâmetros: tamanho dos ramos na filogenia, taxa de transição/transversão (κ) e frequência de bases na terceira posição do códon. Máxima verossimilhança foi utilizado para estimar os parâmetros dos modelos de substituição de códons e na comparação dos modelos. Sítios sob seleção positiva foram identificados utilizando a abordagem Bayes Empirical Bayes (BEB) que calcula a probabilidade posterior das classes dos sítios (Yang et al., 2005). Os modelos descritos acima foram implementados por Nielsen e Yang (1998) e disponíveis através do programa CODEML do pacote PAML (Yang et al., 2000).

Todas as simulações foram no rodas Simulações foram rodadas no servidor altix detalhes do processador do laboratório de Biologia e Diversidade Molecular da Universidade Federal do Rio de Janeiro. Detalhes sobre todos os programas utilizados no estudo encontram-se no anexo F.

Capítulo 4

Resultados

4.1 Estimativas de Taxa Substituição e TMRCA

Os resultados das taxas encontradas para substituição de nucleotídeos e TMRCA para os sorotipos analisados em conjunto e separadamente incluindo linhagens silvestres e endêmico/epidêmicas são apresentados na Tabela 4.1. A convergência do tamanho efetivo mínimo *Effective Sample Sizes* (ESSs) foram maiores que 250 em todas as análises e os valores de 95% de HDP correspondem a probabilidade posterior.

Tabela 4.1: Estimativas de substituição de nucleotídeos

Sorotipo	TMRCA em anos (95% HPD)	Tx subs./sítio/ano x 10⁻⁴	ESS taxas
Denv-I	102.88 (88.58-119.17)	7.48 (6.61-8.43)	258.7
Denv-II	338.08 (227.11-462.97)	7.56 (6.05-8.60)	321
Denv-III	108.93 (87.47-132.62)	8.17 (7.21-9.16)	531.7
Denv-VI	203.35 (116.37-295.74)	7.82 (6.32-9.47)	592
Todos	1672 (1158-2294)	7.64 (6.57-8.74)	

Os resultados mostram que não houve variação significativa nas taxas de substituição de nucleotídeos entre os sorotipos variando pouco entre eles (Figura 4.1). Taxas semelhantes são encontradas na análise conjunta das taxas ($7.64 * 10^{-4}$ subs./sítio/ano).

O tempo de ancestral comum das amostras foi consideravelmente diferente

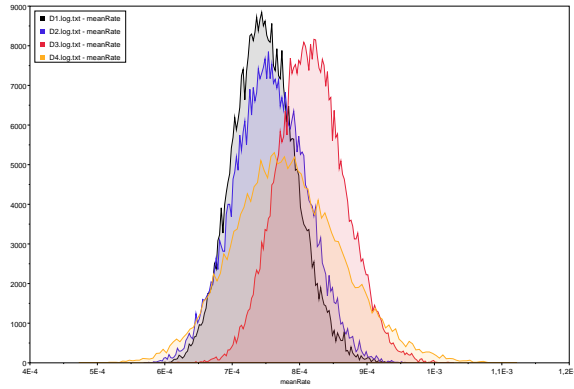


Figura 4.1: Taxas de substituição de nucleotídeos para as amostras analisadas.

nos sorotipos. Denv-II apresentou um maior tempo de divergência das sequências (338 anos) ou seja, divergem por volta de 1670. Os sorotipos Denv-I e III apresentaram um tempo de coalescência mais recente entre 1906 e 1900 respectivamente. A análise conjunta mostrou que todos os sorotipos coalescem aproximadamente desde 336 DC embora não seja possível estimar o intervalo de confiança neste caso.

O coeficiente de variação do relógio molecular teve uma media aproximada de 0.35 para Denv-I e Denv-III e 0.43 para Denv-II e IV (Figura 4.2).

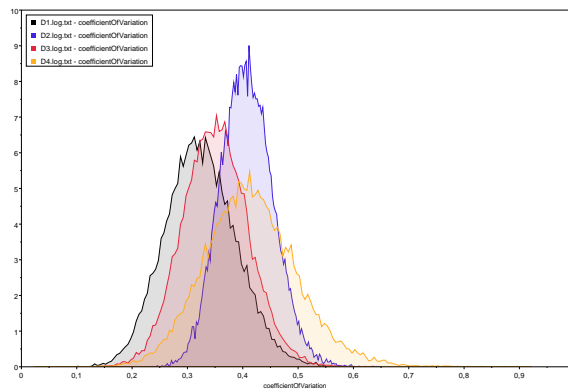


Figura 4.2: Coeficiente de variação do relógio molecular relaxado.

4.2 Genalogia, Padrões de distrições, Diversidade intrasorotípica e Dinâmicas populacionais

Para facilitar o estudo dos padrões geográficos observados, as árvores genealógicas mostradas a seguir tiveram os países coloridos de acordo com a região geográfica. Assim, amostras americanas ficaram em azul, asiáticas em vermelho, africanas de verde e proveniente da Oceania em amarelo.

Desta forma, foram encontrados padrões geográficos em todos os sorotipos analisados. A estrutura intrasorotípica mostrou estrutura geográfica mais restrita em algumas linhagens e estruturas mais cosmopolita em outras. A genealogia sugere que as linhagens silvestres encontradas em Denv-I, II e IV emergiram separadamente das linhagens endêmico/epidêmicas.

A análise conjunta dos sorotipos mostrou Denv-IV como o sorotipo mais divergente seguido de Denv-II. Denv-I e Denv-III foram os sorotipos mais relacionados geneticamente.

A dinâmica populacional do BSP apresentou padrões distintos nos quatro sorotipos analisados. O tempo foi medido em unidades de mutações por sítio e as estimativas medianas são mostradas com linhas sólidas grossas e os limites HPD 95% estão mostrados pelas áreas sombreadas de cinza.

4.2.1 Dengue I

Das 143 sequências de Denv-I foram estabelecidos 5 genótipos distintos (Figura 4.7).

O genótipo I corresponde amostras principalmente sudeste asiático (Tailândia, Mianmar, Vietname, Cambodja, Cingapura), africanas recentes (Arábia Saudita 2004, número de acesso: AM746216; 2005, acesso: AM746214; 2006, acesso: AM746213) e linhagens aparentemente extintas do subgrupo das Américas (Havaí de 1945, acessos: EU848545, AF425619, X76219).

O genótipo II foi caracterizado como cosmopolita e apresentou linhagens asiáticas do sudeste e extremo oriente como Japão, Indonésia, Filipinas, China

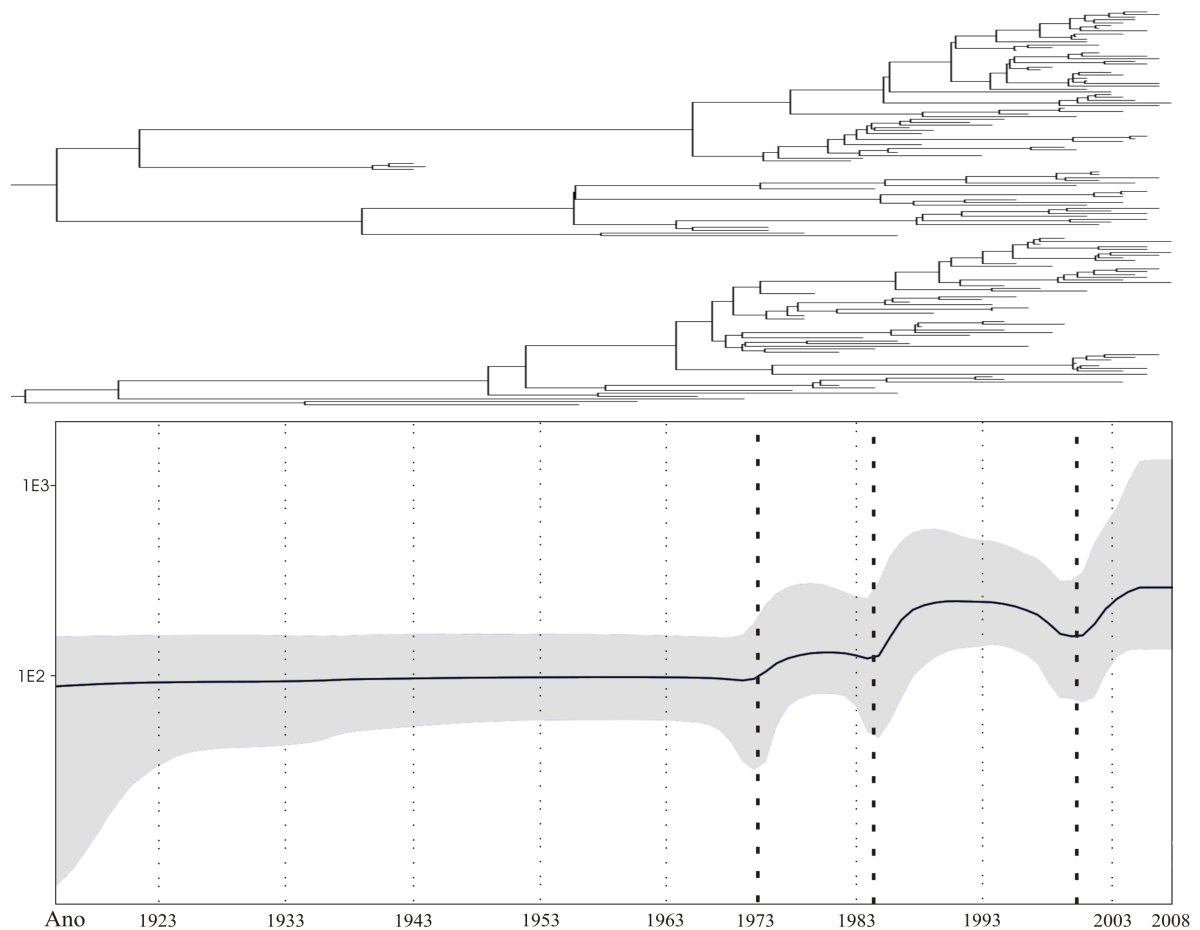


Figura 4.3: Dinâmica populacional em Denv-I. A região sombreada representa 95% HDP e a linha sólida a estimativa média.

e Malásia, africanas com introdução recente (Seicheles 2004 acesso: DQ285561; Madagascar 2006, acesso: EU448412; Reunião 2004, acesso: EU 282328), Oceania (Austrália e Polinésia Francesa) e subgrupo americano aparentemente extinto (Jamaica 1977, acesso: D00501 e Aruba 1985, acesso: D00505).

O genótipo determinado como III é composto principalmente de amostras americanas como Brasil, Venezuela, México, Colômbia, Porto Rico e Nicarágua, recente introdução na Índia 2006, acesso: EU448413, africanas como Costa do Marfim, Comodoros, Angola, Nigéria e Arábia Saudita e amostras asiáticas como Tailândia, Índia e Mianmar, e Coréia do Sul.

O genótipo IV apresenta apenas uma sequência da Malásia 1972, acesso: AF231721 de origem silvestre. O genótipo V também está pouco representativo

contendo apenas duas linhagens da Tailândia sendo a primeira de 1958, acesso: D10513 e a segunda de 1963, acesso: AF425629. Os dois últimos genótipos estabelecidos são amostras antigas e podem estar extintas não correspondendo ao padrão observado atualmente.

O padrão filodinâmico de Denv-I apresentou estrutura geográfica porém estrutura temporal fraca.

Em todos os genótipos foi observada a presença de amostras asiáticas principalmente da região do sudeste asiático. As amostras da Oceania e americanas (cosmopolitas) estão relacionadas a apenas um genótipo. Apesar do pequeno número de amostras, o continente africano se relacionou com três genótipos nas quais dois são introduções aparentemente recentes de provável origem asiática.

Denv-I apresentou um padrão de crescimento de tendência crescente e oscilante. O primeiro crescimento aparece a partir de 1973 seguido de uma estabilização até 1985 com tendência a um padrão exponencial até 1988 com ligeiro declínio até 2000 (Figura 4.3).

4.2.2 Dengue II

Denv-II foi o sorotipo com um maior número de amostras (207 sequências). Foram determinados 5 genótipos distintos com forte estrutura temporal e espacial encontrada (Figura 4.8).

O genótipo denominado I contém linhagens americanas principalmente da América do Sul (Brasil, Porto Rico, Colômbia, Venezuela) e Central (Nicarágua) e um subgrupo com amostras do sudeste asiático (Vietname, Cambodja, Tailândia, Taiwan e Mianmar). O genótipo II é representado por amostras asiáticas com prevalência tailandesa mostrando uma forte estrutura temporal, um subgrupo que contém linhagens colombianas (1944 acessos: FJ390389 e EU854293) aparentemente extintas, linhagens provenientes da Oceania (Papua Nova Guiné 2008, acesso: FJ906959), Filipinas e Tailândia e uma única amostra americana El Salvador de 2000, acesso: DQ364560.

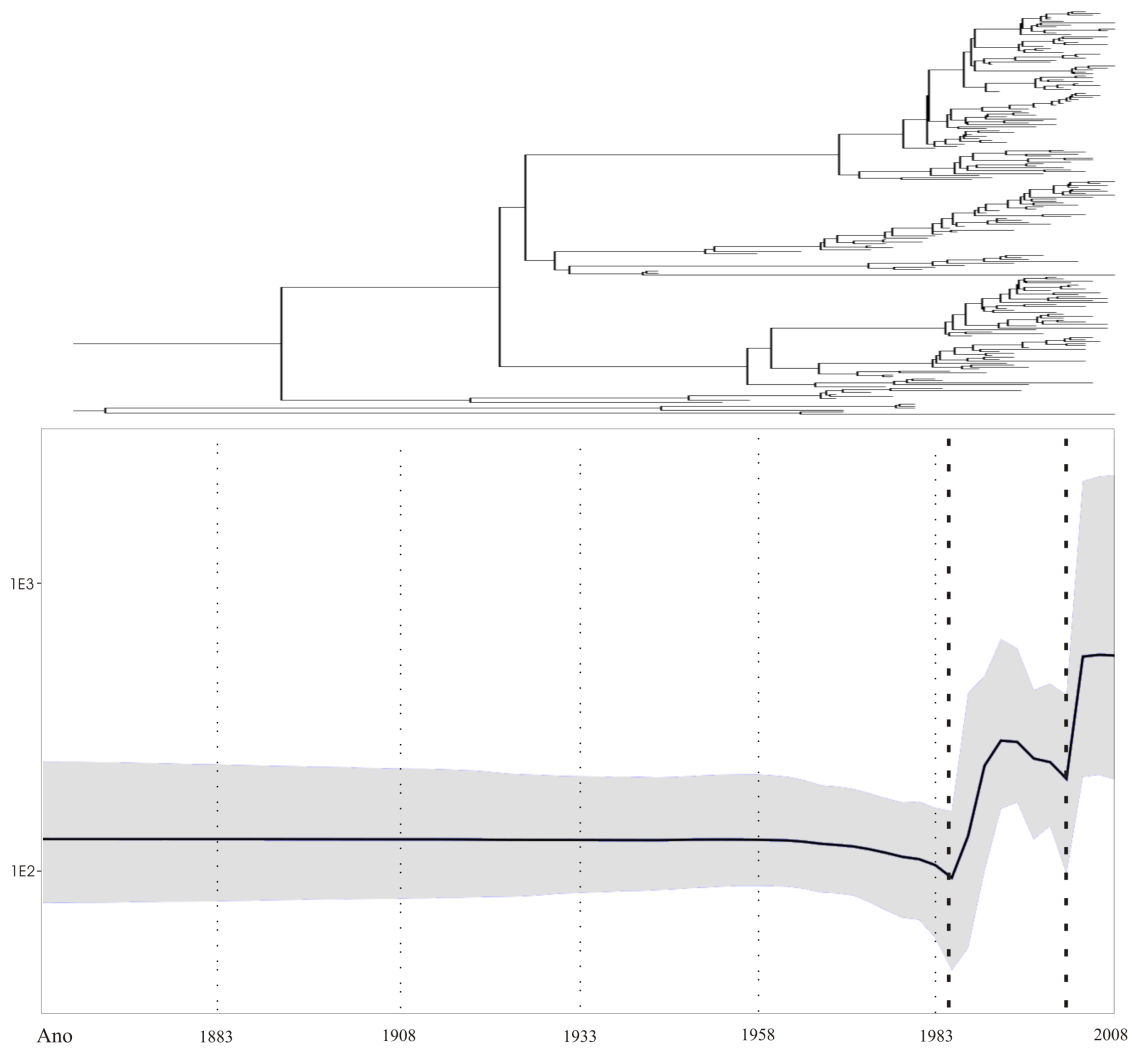


Figura 4.4: Dinâmica populacional em Denv-II. A região sombreada representa 95% HDP e a linha sólida a estimativa média.

O genótipo III foi classificado cosmopolita com linhagens do pacífico como Austrália e Ilhas Cook, africanas como Arábia Saudita, Seicheles, Ghana, Somália e Uganda e duas sequências americanas do México (1988) e Suriname (1996). Este genótipo não apresentou amostras recentes com variação temporal de 1953 a 1995.

O genótipo IV apresentou amostras americanas como Honduras, Peru, Porto Rico, pacífico (Tonga, 1974) e asiático (Índia, 1957).

Por último, o genótipo VII foi caracterizado por amostras silvestres asiáticas (Malásia 1970 e 2008, acessos: FJ467493, AF231717 respectivamente).

Foi observado presença de amostras asiáticas em todos os genótipos e uma

baixa prevalência de amostras de origem da Oceania. As amostras africanas ficaram relacionadas a linhagens cosmopolitas e silvestres. Já as americanas formaram um grande grupo com amostras desde 1983 e outros mais dispersos com amostras menos recentes.

O padrão observado para Denv-II parece indicar infecções de baixo nível endêmico para um rápido crescimento exponencial após 1980. O primeiro pico de crescimento exponencial aparece por volta de 1988 e continuou até 1993 seguido por um declínio durante cinco anos. O segundo pico surge por volta de 2000 com estabilização até o presente (Figura 4.4).

4.2.3 Dengue III

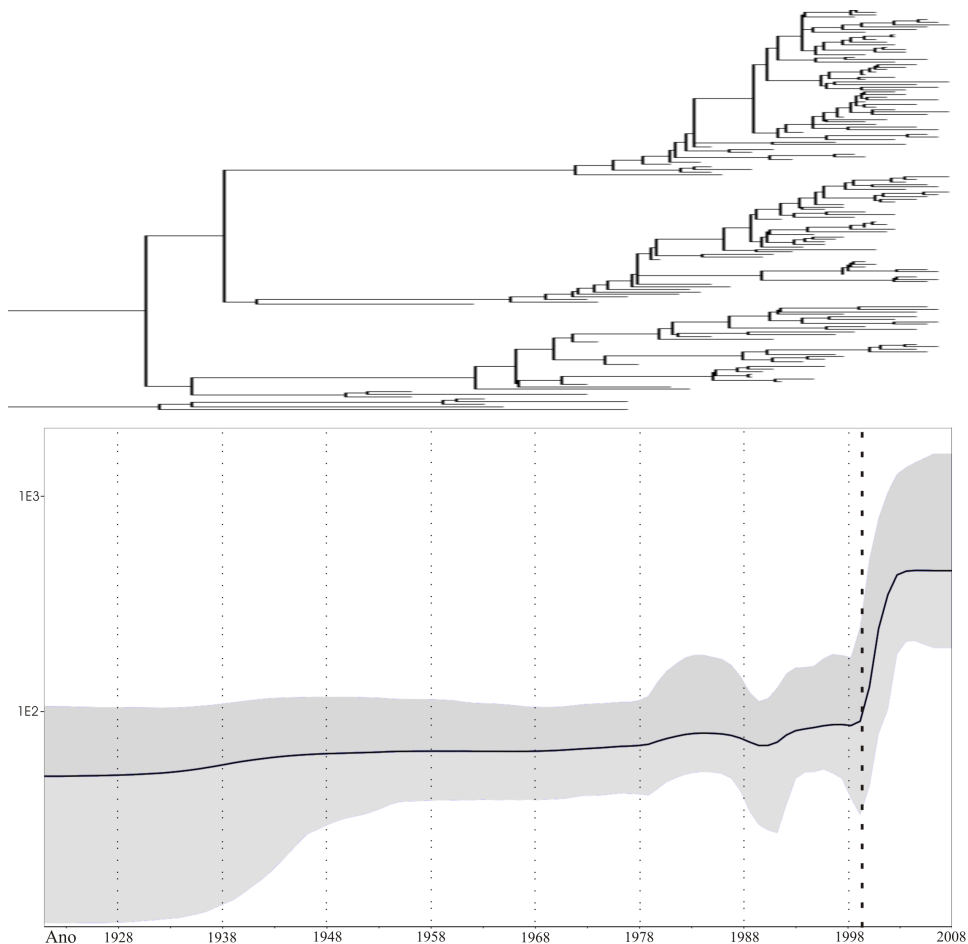


Figura 4.5: Dinâmica populacional em Denv-III. A região sombreada representa 95% HDP e a linha sólida a estimativa média.

Das 163 amostras de Denv-III foram encontrados 4 genótipos distintos (Figura 4.9) com estrutura espacial e temporal.

O genótipo I corresponde a amostras cosmopolita com predominância da América do Sul (Venezuela, Colômbia, Brasil e Suriname) e Central (Porto Rico, México). Também foram encontradas amostras da Ásia meridional (Sri Lanka e Índia), africanas (Arábia Saudita, Moçambique) e apenas uma oceânica (Samoa). O genótipo II é representado totalmente por amostras asiáticas provenientes da Ásia meridional e sudeste (Filipinas, Cambodja, Vietname, Mianmar, Bangladeche, Tailândia, Índia e Malásia).

O genótipo III representa linhagens da Ásia meridional, sudeste e extremo oeste (Índia, Filipinas, Malásia, Taiwan e Japão) e subgrupo da Oceania (Polinésia Francesa, Ilhas Cook, Samoa, Fiji e Nova Caledônia) com provável introdução neste continente originado da Indonésia.

O genótipo IV apresenta amostras americanas (Porto Rico 1963, 1977 e Venezuela 1963, 1977) e oceânicas (Polinésia Francesa, 1965, acesso: L11439) aparentemente extintos.

Amostras asiáticas apareceram em três genótipos. Sequências da Oceania datam de 1986 a 1996. Já as amostras americanas corresponderam a dois genótipos distintos sendo um grande com amostras circulantes desde 1985 e outro aparentemente extinto. Para África há uma baixa prevalência de amostras contidas dentro do genótipo I de provável origem asiática.

Denv-III indicou um crescimento constante até 2000 seguido de um acentuado crescimento semelhante a um padrão exponencial durante dois anos. Logo após ocorreu estabilização até o presente (Figura 4.5).

4.2.4 Dengue IV

Denv-IV foi o grupo com menor representatividade de amostras (90 sequências). Entretanto apresentou diversidade genotípica com quatro grupos distintos (Figura 4.10).

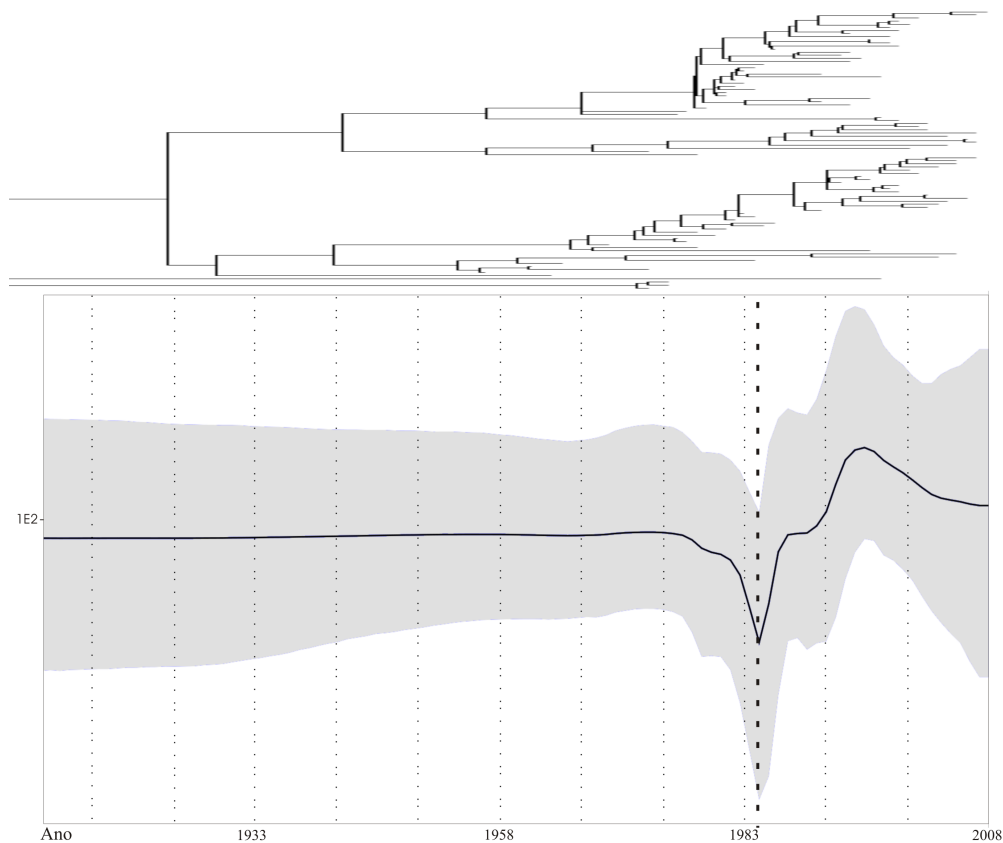


Figura 4.6: Dinâmica populacional em Denv-IV. A região sombreada representa 95% HDP e a linha sólida a estimativa média.

O genótipo I corresponde amostras cosmopolitas provenientes das Américas (Venezuela, Porto Rico, Barbados, México, El Salvador, Jamaica), amostras da Oceania como Polinésia Francesa (1979 e 1985), Ilhas Salomão e Nova Caledônia (1984) com origem americana, amostras asiáticas da Índia (1977), Malásia, Indonésia, Malásia, Japão e Timor Leste.

O genótipo II foi representado por amostras asiáticas principalmente tailandesas com estrutura espacial forte. Além da Tailândia, também apresentou amostras provindas do Vietname, Filipinas, Japão e Sri Lanka. O genótipo III é representado por uma única amostra da Tailândia (1997, acesso: AY618979). O genótipo V representa linhagem silvestre da Malásia 1973, (acesso: AF231724) e 1975, (acessos: AF231723, AF231725).

O sorotipo Denv-IV não apresentou nenhuma amostra de origem africana.

O continente asiático apareceu em todos os genótipos. As poucas amostras da Oceania constaram em dois sorótipos. Já as americanas ficaram restritas e têm uma provável origem asiática.

Em Denv-IV houve um padrão de crescimento constante até 1974 seguido de um declínio acentuado até 1986. Logo após ocorre um crescimento também acentuado seguido de declínio por volta de 1995 até o presente sugerindo um efeito gargalo da população viral durante 10 anos. A recuperação após este efeito parece ser rápida e exponencial até 1995 (Figura 4.6).

4.3 Índícios de Sítios sob Pressão Seletiva

Não foram encontradas evidências fortes ($IC > 0.98$) de seleção positiva atuando em aminoácidos do gene do ENV em nenhum modelo e sorótipo analisado Tabelas 4.2, 4.3, 4.4.

Os valores da taxa de transição/transversão (κ) apresentaram valores maiores em Denv-IV ($\kappa M2 = 14.1027$) e menores em Denv-II ($\kappa M2 = 9.1735$).

No modelo M2, a análise BEB sugeriu alguns sítios na qual $\omega > 1$ para os sorótipos Denv-I, III, IV, indicando seleção positiva (Figura 4.3) entretanto com IC baixo (tabela).

Tabela 4.2: Resíduos sob pressão seletiva no modelo M0

sorótipo	l	κ	ω
Denv-I	-11888.9486	9.70633	0.06753
Denv-II	-15104.470	5.39708	0.03487
Denv-III	-10588.5747	11.7601	0.07183
Denv-IV	-8439.3015	13.41567	0.05449

Tabela 4.3: Resíduos sob pressão seletiva no modelo M1

sorotipo	l	κ	p_0	p_1	w_0	w_1
Denv-I	-11831.140	10.08117	0.94275	0.05725	0.04225	1.00000
Denv-II	-14997.5132	9.12979	0.97409	0.02591	0.03359	1.00000
Denv-III	-10483.4030	11.96147	0.94908	0.05092	0.03831	1.00000
Denv-IV	-8333.7612	14.10257	0.93908	0.06092	0.02080	1.00000

Tabela 4.4: Resíduos sob pressão seletiva no modelo M2

sorotipo	l	κ	p_0	p_1	p_2	w_0	w_1	w_2
Denv-I	-11831.141	10.0701	0.94275	0.05725	0.00000	0.04226	1.000	24.144
Denv-II	-14997.787	9.173	0.97392	0.02552	0.00055	0.03360	1.000	10.634
Denv-III	-10483.607	11.965	0.94907	0.05093	0.00000	0.03831	1.000	44.309
Denv-IV	-8333.761	14.103	0.93908	0.06092	0.00000	0.02080	1.000	30.016

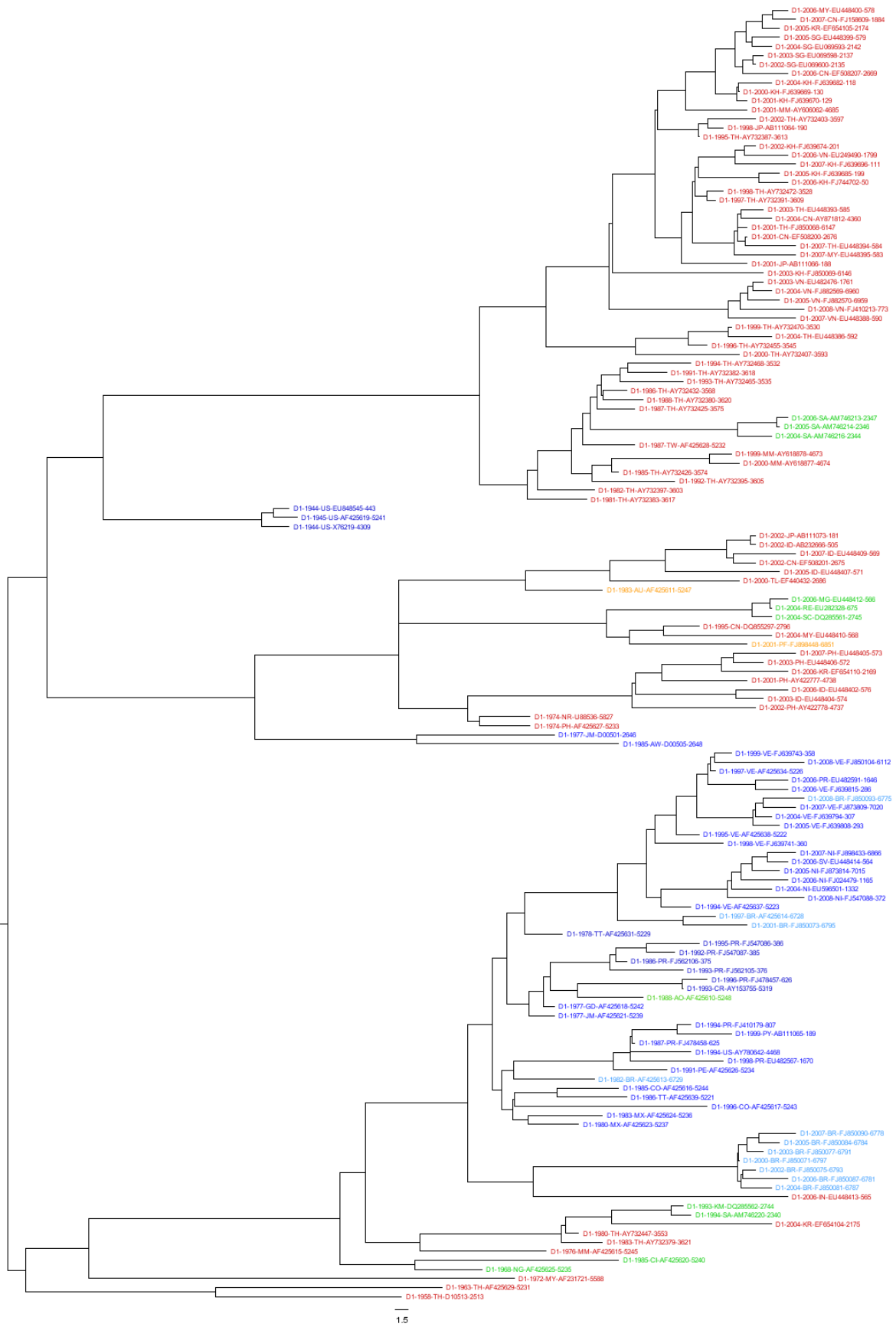


Figura 4.7: Árvore Filogenética para Denv-I.

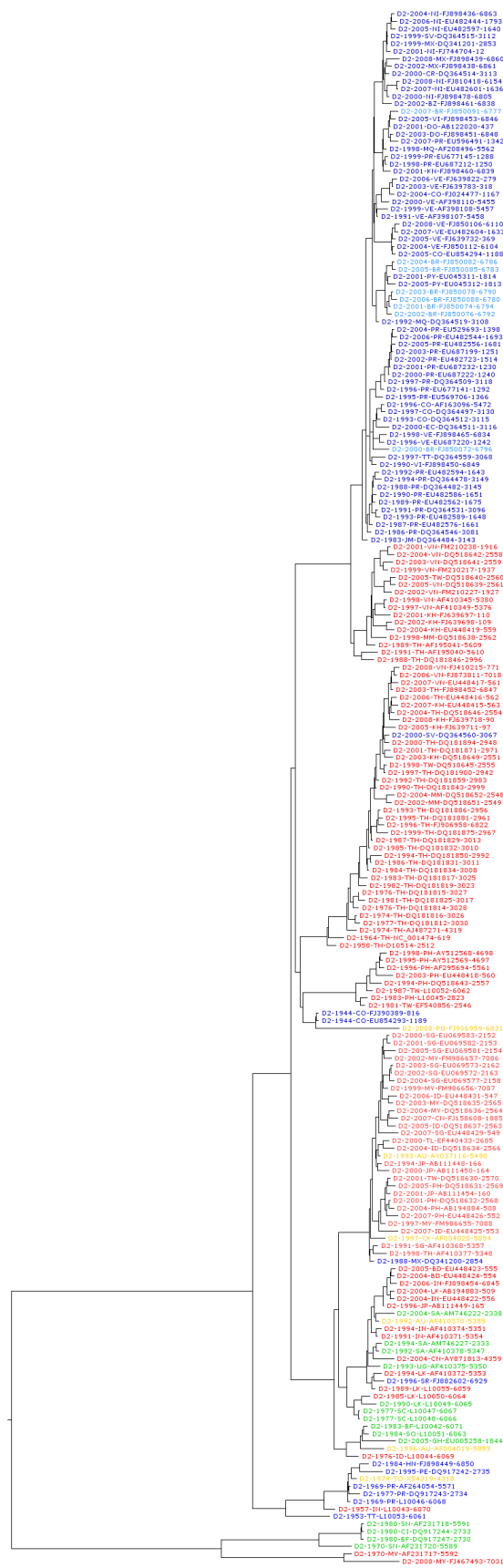


Figura 4.8: Árvore Filogenética para DENV-II.

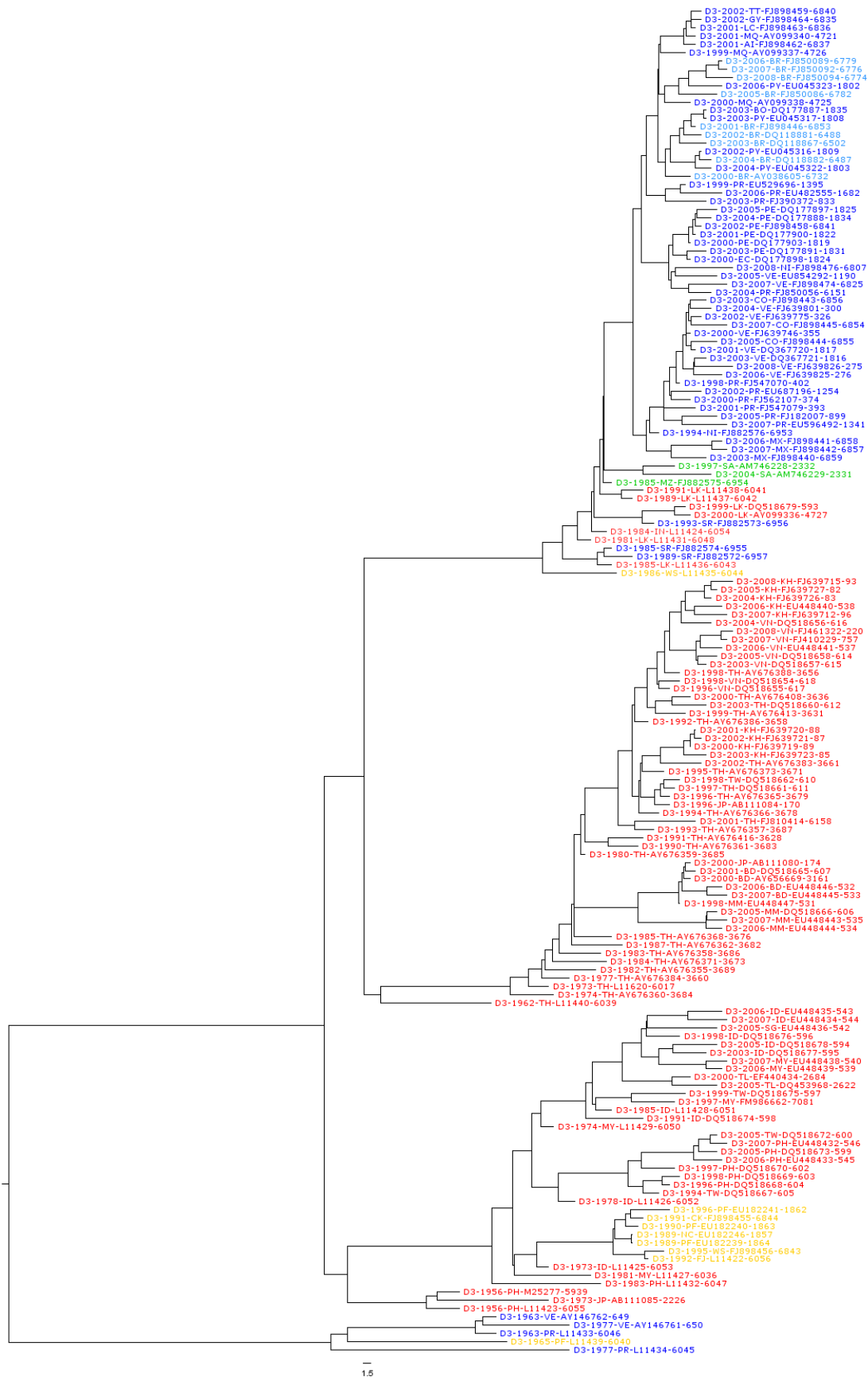


Figura 4.9: Árvore Filogenética para Denv-II.

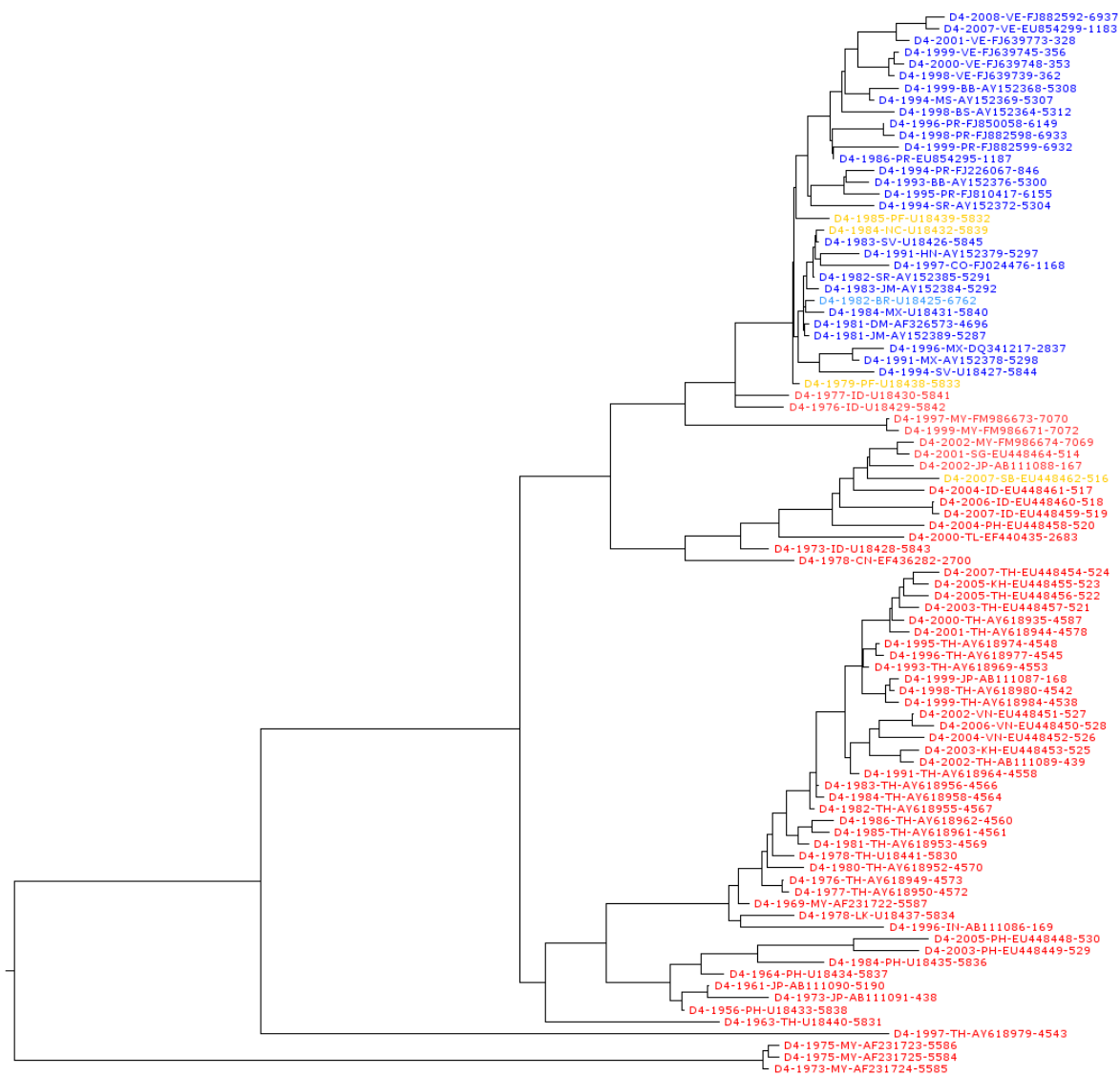


Figura 4.10: Árvore Filogenética para Denv-IV.

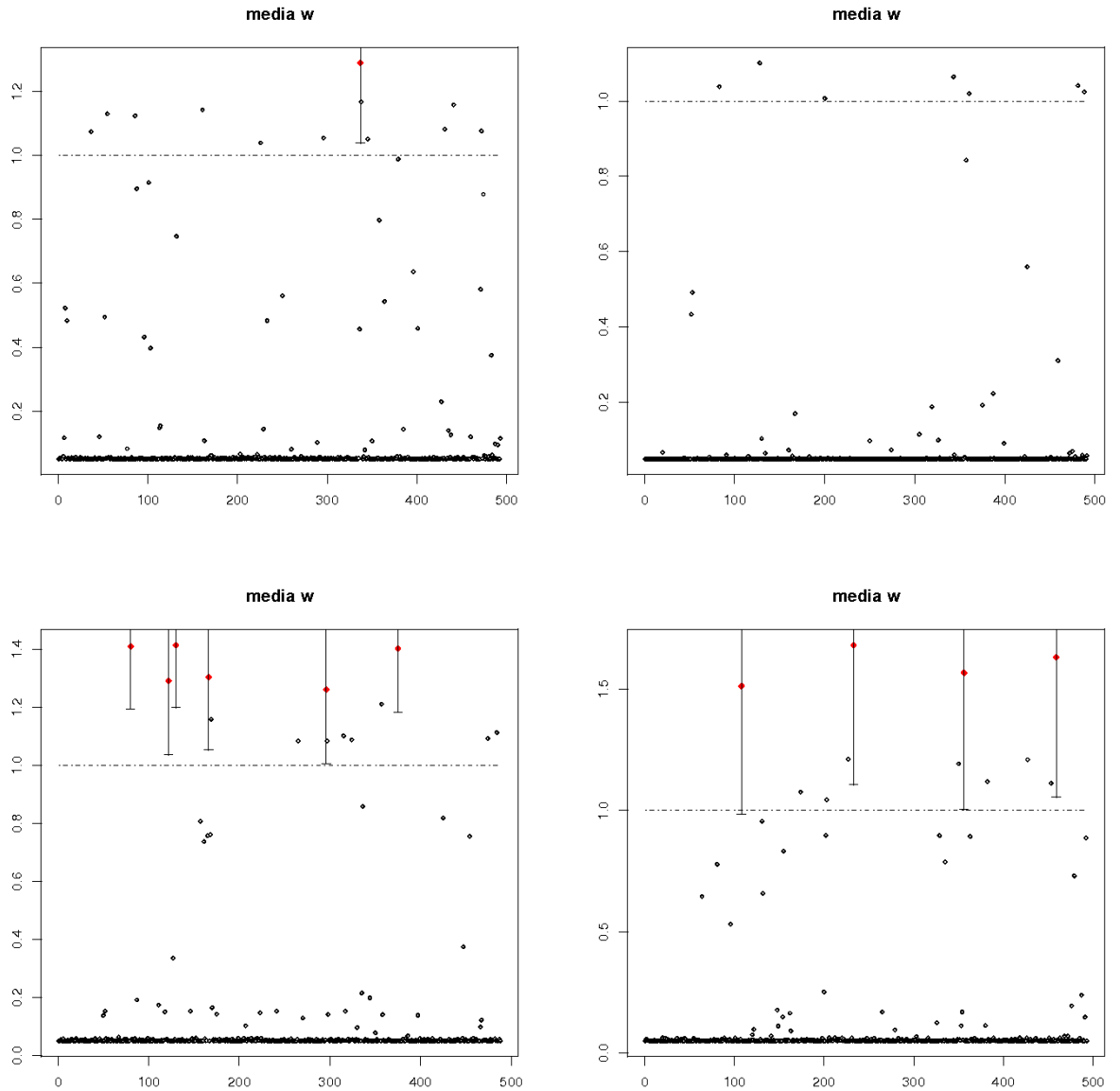


Figura 4.11: Valores de ω para DenvI-IV respectivamente. Os pontos plotados em vermelhos indicam valores de ω com probabilidade posterior maior que 1 com intervalo de confiança.

Capítulo 5

Discussão

Taxas evolutivas

As taxas de substituição de nucleotídeos observadas neste estudo ficaram dentro das taxas medias estimada para vírus de RNA (Twiddy et al., 2003a).

Neste estudo, não houve variação significativa entre os sorotipos e foram similares aos trabalhos de Jenkins et al. (2002); Zanotto et al. (1996); Carrington et al. (2005).

Os estudos conduzidos por Twiddy et al. (2003a) e (Klungthong 2001) mostraram diferenças nas taxas de substituição de nucleotídeos entre os sorotipos variando entre 4.55×10^{-4} subst./sítio/ano em Denv-I a 10.72×10^{-4} subst./sítio/ano para Denv-IV. Novamente as diferenças podem estar ligadas ao tamanho da amostra considerada e a retirada de sequências silvestres.

Taxas de substituição de nucleotídeos são cruciais para aumentar a informação sobre a história epidemiológica e compreensão dos processos que controlam a evolução viral e consequentemente predizerem respostas de tratamentos com drogas ou programas de vacinação (Holmes e Burch, 2000). Variações em taxas virais observadas devem-se a erros provenientes da RNA polimerase, tamanho do genoma considerado e velocidade de replicação. No geral, viroses que são transmitidas por mosquitos vetores apresentam uma taxa ligeiramente mais lenta que viroses sem a presença de vetores como influenza e HIV como demonstrado por Zanotto et al. (1996) e (Jenkins et al., 2002). De acordo

com os resultados obtidos as taxas estimadas não seguiram o relógio molecular em nenhum sorotipo condizente com Twiddy et al. (2003a).

Seleção

É de se esperar que taxas de substituição não sinônimas variem amplamente entre os vírus de RNA (e entre os genes virais) e essas diferenças têm apenas um pequeno impacto sobre as taxas de substituição total, embora resultem em mudanças nas sequências de aminoácidos da proteína. Em arboviroses, a pressão de seleção purificadora observada é maior que viroses com apenas um hospedeiro. Tal fato observado está associado com a replicação em diferentes espécies hospedeiras Weaver e Barrett (2004); Jenkins et al. (2002) como observado em flavivirus (Holmes, 2003; Klungthong et al., 2004).

Confirmando a hipótese anterior quando não houve evidências de pressão seletiva atuando no gene ENV pelos métodos de análise de seleção. Os sorotipos Denv-I, III e IV indicaram sítios com w maior que 1, ou seja, indicativo de seleção positiva, mas com intervalo de confiança abaixo de 95. Klungthong et al. (2004) não encontrou indícios de evolução adaptativa em nenhum códon ou linhagem de Denv-IV.

A falta de evidência de seleção positiva ocorrendo no envelope viral, região importante no reconhecimento da célula hospedeira pode estar relacionada a análises de padrões gerais. Estudos de genótipos específicos relacionados com dengue hemorrágica apresentaram regiões do ENV sob seleção positiva no sorotipo Denv-II, conforme demonstrado por (Bennett et al., 2006) e (Twiddy et al., 2002) avaliando pressões seletivas em genótipos de Denv-II.

Filodinâmicas, Diversidade e Linhagens extintas e emergentes

As 603 sequências da região do ENV em Denv-I a IV revelaram uma forte estrutura espacial com subdivisão das linhagens virais. Os grandes agrupamentos geográficos compreenderam as Américas (Central e Sul) e principalmente a Ásia

(Sudeste Asiático e Meridional) coerentes com trabalhos apresentados por (Araujo et al., 2009; Holmes, 2004).

Outros padrões de distribuição geográfica observados parecem revelados em uma escala mais fina. Por exemplo, o padrão satélite central e onda foram observados no sudeste asiático especificamente na Tailândia região que parece sustentar epidemias e reservatórios fornecendo cepas aos países vizinhos Camboja e Vietname, tal observação foi comum a todos os sorotipos.

Na transmissão em onda a virose se difunde de uma região para regiões vizinhas sendo difícil de observar em uma escala global para dengue, pois a transmissão também ocorre devido ao intenso fluxo de pessoas em regiões intercontinentais favorecido pelo transporte aéreo (Weaver e Vasilakis, 2009).

A estrutura temporal é causada pelo surgimento contínuo e extinção de linhagens virais. As novas linhagens são regularmente produzidas por mutação, mas a maioria extinta em períodos relativamente rápidos e com progresso em subseqüentes amostragens (Klungthong et al., 2004). A estrutura temporal foi notada nos países asiáticos (Cambodia (Denv-I), Tailândia (todas), Vietname (Denv-I,II, III), Bangladesh (Denv-III), Filipinas (Denv-II, III), Mianmar (Denv-III), Singapura (Denv-II), Malásia (Denv-IV), Índia (Denv-IV) e nos países americanos Nicarágua (Denv-I, II), Venezuela (todos), Brasil (Denv-I, II, III), Colômbia (Denv-II), Porto Rico (Denv-II,III, IV), Peru (Denv-III), México (Denv-III).

Nas Américas é bem conhecido que o dengue circulante é de origem asiática (Aquino et al., 2009; Lanciotti et al., 1997). Em relação aos genótipos americanos atuais, Denv-I genótipo III é o que circula há mais tempo nas Américas com ancestrais desde 1977. Já Denv-III, genótipo I circula mais recentemente nas Américas formando um clado distinto desde 1994 na Nicarágua. Os sorotipos Denv-II genótipo I e Denv-IV genótipo I circulam nas Américas envolvendo ancestrais existentes desde a década de 80. Amostras americanas também parecem influenciar na transmissão para Ásia como foi notada na amostra Índia 2006

(acesso: EU448413) com origem brasileira.

As recentes introduções africanas foram observadas em Denv-I em dois genótipos diferentes. No genótipo I a introdução na Arábia Saudita parece ter origem tailandesa. Já a introdução nas ilhas de Seicheles 2004, Reunion 2004 e em Moçambique 2006 parecem ter origem chinesa. Também foram observadas introduções recentes na Arábia Saudita de Denv-II (2004) e III (2004), indicando a ocorrência de hiperendemicidade neste país.

A quantidade limitada de amostras no continente africano obtida é reflexo da fiscalização limitada ou inexistente neste continente e dificulta o entendimento da circulação do vírus da dengue no continente.

Na Oceania parece circular recentemente apenas Denv-IV encontrada nas Ilhas Salomão em 2007 com provável origem da Indonésia. As demais amostras desta região são antigas e parecem formar clados extintos.

Em todos os sorotipos foi observado clados e linhagens sem presença de amostras recentes (após o ano 2000). Em Denv-I os genótipos I e II Ásia contem amostras americanas do Havaí (1944) e Aruba (1985). Em Denv-II e II os genotipo IV são considerados aparentemente extintos.

Caracterização genotípica é importante no monitoramento de mudanças genéticas aparentes no tempo em DENV e identificar movimentos em fina escala. No geral, a quantidade de genotipos estabelecidos mostrou pouca variação e concordância com trabalhos já realizados por Twiddy et al. (2002); Lanciotti et al. (1994); Holmes (2003); Messer et al. (2003); Bennett et al. (2003)

Diferenças observadas em outros trabalhos de (Goncalvez et al., 2002) para Denv-I Dias, 2006, (Wang et al., 2000; Araujo et al., 2009) devem-se em grande parte, à quantidade amostral de sequências utilizadas bem como as ausência de registros (sequencias) em localidades chaves especificas e tempo das amostras. Dentre estas localidades chaves há ausência de sequências da região do envelope para Denv-II nas Américas durante a década de 80 e Denv-III a partir de 1994 na mesma região. O aumento de sequências pode assim refinar as rotas do vírus em

regiões geográficas e estabelecer padrões filogeográficos mais precisos.

Além disso, as estruturas genotípicas são possíveis de revisão e não seguem necessariamente um único padrão definido. Genealogias mais recentes podem mostrar extinção ou colapso de linhagens não observadas anteriormente além de dependerem da disponibilidade das sequências (Kyle e Harris, 2008).

TMRCA

O tempo de ancestral comum foi variável, nos sorotipos, em parte devido à ocorrência de linhagens silvestres contidas nos sorotipos de Denv-I, II e IV na qual apresentaram ancestrais comuns distintos de linhagens envolvidas no ciclo endêmico/epidêmico. Nos dados conjuntos o tempo de ancestral comum foi por volta de 336 DC coincidente com os primeiros relatos dos sintomas semelhantes a dengue ocorridos na dinastia chinesa de Chin. Os estudos de (Twiddy et al., 2003a) e (Lewis et al., 1993) levantaram um tempo de divergência sorotípica mais recente ocorrendo por volta do ano de 1115 DC.

pergunta: o que levou a diversificação das linhagens em dengue? como e qd aconteceu?

Linhagens silvestres

De acordo com a árvore genealógica e tempo do ancestral comum, as linhagens silvestres emergiram separadamente das linhagens endêmicas/epidêmicas como também apresentado no trabalho de (Wang et al., 2000).

O ciclo silvestre parece ser uma linhagem mais antiga, pois a eficiência de transmissão em humanos requer um tamanho populacional mínimo de 10.000 a 1 milhão de habitantes (Gubler, 1998).

As estimativas para a idade do TMRCA tanto das cepas silvestres quanto as linhagens humanas sugerem o sorotipo Denv-II como o primeiro a desenvolver o ciclo de transmissão sustentada nos seres humanos mais cedo que nos outros sorotipos. Tal evidência também é apontada por (Twiddy et al., 2003b).

Pouco se conhece da transmissão entre linhagens silvestres e humanas. Alguns estudos apresentados por Rudnick (1966) documentam casos históricos de infecções humanas por genótipos silvestres.

Dinâmica populacional e severidade

Padrões de distribuição geográfica são influenciados pela dinâmica populacional. O efeito fundador de linhagens provoca grandes surtos epidêmicos que levam a diversificação e uma aceleração nas taxas de crescimento populacional.

Foi observado que as introduções de Denv-II e IV nas Américas em 1981 coincidem com explosões demográficas observadas no BSP com estágios iniciais de crescimento acelerado. O mesmo não foi visto por Carrington et al. (2005) que encontrou padrões iniciais de invasão e expansão distintos na qual Denv-IV aumenta mais rapidamente que Denv-II. Entretanto a autora avaliou somente genótipos específicos das Américas e valores de taxas de crescimento. Observa-se em Denv-IV um declínio acentuado anterior a introdução nas Américas com um efeito gargalo. A introdução nas Américas deste sorotipo gerou um nicho vago como provavelmente a falta de imunidade a este sorotipo que acarretou acentuado crescimento, diversificação e expansão nas Américas e novamente na Ásia.

As explosivas atividades virais observadas ocorreram em um curto espaço de tempo e em épocas diferentes em cada sorotipo. Posteriormente as explosões é observada uma estabilização curta do crescimento e no caso de Denv-IV novamente um crescimento acentuado. Em Denv-II após um declínio, nova explosão virótica ocorre por volta de 2000 em todos os genótipos atuais.

A introdução de Denv-I nas Américas em 1977 acompanha um ligeiro aumento no crescimento populacional. A última onda de crescimento observada recentemente acompanha expansão em todos os genótipos. É conhecido que Denv-I relaciona-se com grandes epidemias na região meridional asiática, como na Índia na qual é notada a presença há mais de 50 anos (Kukreti et al., 2009) circulando juntamente com Denv-II e III.

No geral, os declínios e estabilizações observadas podem estar relacionados com fatores de neutralização viral por imunidade cruzada entre linhagens proximamente relacionadas em uma população exposta Araujo et al. (2009).

Além disso, a limitação do vetor deve ser levada em consideração nos padrões demográficos observados em arbovirose. Sob condições de epidemia, cada indivíduo infectado tem uma probabilidade de transmitir o vírus para vários contatos, assim, aumentar o nível geral de diversidade da população. Após um período de epidemia de transmissão sustentada, espera-se uma diversidade maior do que seria o caso em condições endêmicas, onde as taxas de transmissão são mais baixas (Twiddy et al., 2003a). No caso da dengue, o inseto *Aedes* tem seu ciclo de vida altamente dependente de água limpa e parada. Assim as taxas de transmissão variam dramaticamente entre chuvosa e seca como resultados da disponibilidade de locais reprodutivos para o mosquito vetor. Adicionalmente, a circulação de sorotipos distintos em uma mesma região geográfica pode levar a padrões complexos de competição seletiva (Zhang et al., 2005). Por exemplo, análises filogenéticas revelaram que uma alta taxa relativa de extinção de linhagens e substituição (reposição) de outros sorotipos geram períodos de flutuações na diversidade que podem levar ao maior efeito da estrutura populacional. A estrutura genética populacional de dengue parece ser fortemente influenciada pela extensão da reação cruzada entre sorotipos por reforço e proteção.

A prevalência de cada sorotipo de dengue em uma mesma região tem aumentado o risco de DHF nos últimos anos acompanhado pelo aumento da diversidade genética dentro de cada sorotipo (Twiddy et al., 2003a). Embora nenhuma hipótese de que a Denv-II é diretamente responsável por alguns casos de DHF nas Américas tenha sido testada em algum modelo animal, vários estudos apontam Denv-II como o mais associado com a DHF principalmente nas Américas, seguido de Denv1 e 3 Carrington et al. (2005); RicoHesse et al. (1997). Tais comprovações estão associadas a grande epidemia ocorrida em 1981 em Cuba após introdução do novo genotipo de Denv-II acarretando 350.000 casos sendo 10.000

DHF e 158 óbitos (Kouri, 1986).

Dificuldades encontradas

gaps na amostragem

Capítulo 6

Conclusão & Perspectivas

monitoramento saude publica ...

Referências Bibliográficas

- A. A-Nuegoonpipat, A. Berlioz-Arthaud, V. Chow, T. Endy, K. Lowry, L. Q. Mai, T. U. Ninh, A. Pyke, M. Reid, J. M. Reynes, S. T. S. Yun, H. M. Thu, S. S. Wong, E. C. Holmes, e J. Aaskov. Sustained transmission of dengue virus type 1 in the pacific due to repeated introductions of different asian strains. **Virology**, 329(2):505–512, 2004.
- S. F. Altschul, W. Gish, W. Miller, E. W. Meyers, e D. J. Lipman. Basic Local Alignment Search Tool. **Journal of Molecular Biology**, 215(3):403–410, 1990.
- V. H. Aquino, A. A. Amarilla, H. L. Alfonso, W. C. Batista, e L. T. M. Figueiredo. New genotype of dengue type 3 virus circulating in brazil and colombia showed a close relationship to old asian viruses. **Plos One**, 4(10), 2009.
- J. M. G. Araujo, R. M. R. Nogueira, H. G. Schatzmayr, P. M. D. Zanotto, e G. Bello. Phylogeography and evolutionary history of dengue virus type 3. **Infection Genetics And Evolution**, 9(4):716–725, 2009.
- M. A. Beaumont e B. Rannala. The bayesian revolution in genetics. **Nature Reviews Genetics**, 5(4):251–261, 2004.
- S. N. Bennett, E. C. Holmes, M. Chirivella, D. M. Rodriguez, M. Beltran, V. Vorndam, D. J. Gubler, e W. O. McMillan. Selection-driven evolution of emergent dengue virus. **Molecular Biology And Evolution**, 20(10):1650–1658, 2003.
- S. N. Bennett, E. C. Holmes, M. Chirivella, D. M. Rodriguez, M. Beltran, V. Vorndam, D. J. Gubler, e W. O. McMillan. Molecular evolution of dengue 2

- virus in puerto rico: positive selection in the viral envelope accompanies clade reintroduction. **Journal Of General Virology**, 87:885–893, 2006.
- G. Bian, Y. Xu, Lu.P., Y. Xie, e Z Xi. The endosymbiotic bacterium *Wolbachia* induces resistance to dengue virus in *Aedes aegypti*. **Plos Pathogen**, 6(4):1–10, 2010.
- M. Breitbart e F Rohwer. Here a virus, there a virus, everywhere the same virus? **Trends in Microbiology**, 13(6):278–284, 2005.
- C.V.F. Carrington, J.E. Foster, O.G. Pybus, S.N. Bennett, e E.C. Holmes. Invasion and maintenance of dengue virus type 2 and type 4 in the Americas. **Journal of Virology**, 79(23):14680–14687, DEC 2005. ISSN 0022-538X.
- T. J. Chambers, C. S. Han, R. Galler, e C. M. Rice. Flavivirus genome organization, expression, and replication. **Annual Review Of Microbiology**, 44:649–688, 1990.
- N.J Dimmock, A.J. Easton, e K. Leppard. **Introduction to Modern Virology**. Blackwell, 2007.
- A. J. Drummond, S. Y. W. Ho, M. J. Phillips, e A. Rambaut. Relaxed phylogenetics and dating with confidence. **Plos Biology**, 4:699–710, 2006.
- A. J. Drummond, O. G. Pybus, A. Rambaut, R. Forsberg, e A. G. Rodrigo. Measurably evolving populations. **Trends In Ecology & Evolution**, 18(9): 481–488, 2003.
- A. J. Drummond e A. Rambaut. Beast: Bayesian evolutionary analysis by sampling trees. **Bmc Evolutionary Biology**, 7, 2007.
- A. J. Drummond, A. Rambaut, B. Shapiro, e O. G. Pybus. Bayesian coalescent inference of past population dynamics from molecular sequences. **Molecular Biology And Evolution**, 22(5):1185–1192, 2005.

- L. T. M. Figueiredo. **Informes em Saúde Pública 2: Simpósio sobre Dengue.** Capítulo Patogenia do dengue, páginas 13–24, Universidade De São Paulo, 1999.
- S.J. Flint, L.W. Enquist, V.R. Racaniello, e A.M. Skalka. **Principles of Virology: Molecular Biology, Pathogenesis, and Control of Animal Viruses,** ASM Press, 2004.
- A. P. Goncalvez, A. A. Escalante, F.H. Pujol, J.E. Ludert, D. Tovar, R.A. Salas, e F. Liprandi. Diversity and evolution of the envelope gene of dengue virus type 1. **Virology**, 303(1):110–119, 2002.
- B. T. Grenfell, O. N. Bjornstad, e J. Kappey. Travelling waves and spatial hierarchies in measles epidemics. **Nature**, 414(6865):716–723, 2001.
- B. T. Grenfell, O. G. Pybus, J. R. Gog, J. L. N. Wood, J. M. Daly, J. A. Mumford, e E. C. Holmes. Unifying the epidemiological and evolutionary dynamics of pathogens. **Science**, 303(5656):327–332, 2004.
- D. J. Gubler. Dengue and dengue hemorrhagic fever. **Clinical Microbiology Reviews**, 11(3):480–, 1998.
- D. J. Gubler. Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. **TRENDS in Microbiology**, 10(2): 100–103, 2002.
- B. Guy e J. W. Almond. Towards a dengue vaccine: Progress to date and remaining challenges. **Comparative Immunology Microbiology And Infectious Diseases**, 31(2-3):239–252, 2008.
- M. G. Guzman, M. Alvarez, R. Rodriguez-Roche, L. Bernardo, T. Montes, S. Vazquez, L. Morier, A. Alvarez, E. A. Gould, G. Kouri, e S. B. Halstead. Neutralizing antibodies after infection with dengue 1 virus. **Emerging Infectious Diseases**, 13(2):282–286, 2007.

- S. B. Halstead, J. A. Suaya, e D. S. Shepard. The burden of dengue infection. **Lancet**, 369(9571):1410–1411, 2007.
- W. M. Hammon, A. Rudnick, e G. E. Sather. Viruses associated with epidemic hemorrhagic fevers of the philippines and thailand. **Science**, 131(3407):1102–1103, 1960.
- E. C. Holmes. Molecular clocks and the puzzle of RNA virus origins. **Journal Of Virology**, 77(7):3893–3897, 2003.
- E. C. Holmes. The phylogeography of human viruses. **Molecular Ecology**, 13(4):745–756, 2004.
- E. C. Holmes. Evolutionary history and phylogeography of human viruses. **Annual Review of Microbiology**, 62:307–328, 2008. ISSN 0066-4227.
- E. C. Holmes e S. S. Burch. The causes and consequences of genetic variation in dengue virus. **Trends In Microbiology**, 8(2):74–77, 2000.
- R. R. Hudson. Testing the constant-rate neutral allele model with protein-sequence data. **Evolution**, 37(1):203–217, 1983.
- G.M Jenkins, A. Rambaut, O.G Pybus, e E.C. Holmes. Rates of molecular evolution in RNA viruses: A quantitative phylogenetic analysis. **Journal of Molecular Evolution**, 54(2):156–165, FEB 2002. ISSN 0022-2844.
- T. H. Jukes e C. R Cantor. **In Mammalian protein metabolism**. Capítulo Evolution of protein molecules, páginas 121–123, Academic Press, New York., 1969.
- M. Kimura. Evolutionary rate at molecular level. **Nature**, 217(5129):624–, 1968.
- J.F.C. Kingman. On the genealogy of large populations. **J. Appl. Prob.**, 19A: 27–43, 1982.

- C. Klungthong, R. Putnak, M. P. Mammen, T. Li, e C. Zhang. Molecular genotyping of dengue viruses by phylogenetic analysis of the sequences of individual genes. **Journal Of Virological Methods**, 154(1-2):175–181, 2008.
- C. Klungthong, C. L. Zhang, M. P. Mammen, S. Ubol, e E. C. Holmes. The molecular epidemiology of dengue virus serotype 4 in Bangkok, Thailand. **Virology**, 329(1):168–179, 2004.
- H. Kukreti, P. K. Dash, M. Parida, A. Chaudhary, P. Saxena, R. S. Rautela, V. Mittal, M. Chhabra, D. Bhattacharya, S. Lal, P. V. L. Rao, e A. Rai. Phylogenetic studies reveal existence of multiple lineages of a single genotype of Denv-1 (genotype III) in India during 1956-2007. **Virology Journal**, 6, 2009.
- S. Kumar, M. Nei, J. Dudley, e K. Tamura. MEGA: A biologist-centric software for evolutionary analysis of dna and protein sequences. **Briefings In Bioinformatics**, 9(4):299–306, 2008.
- J. L. Kyle e E. Harris. Global spread and persistence of dengue. **Annual Review Of Microbiology**, 62:71–92, 2008.
- R. S. Lanciotti, J. G. Lewis, D. J. Gubler, e D. W. Trent. Molecular evolution and epidemiology of dengue-3 viruses. **Journal Of General Virology**, 75:65–75, 1994.
- R. S. Lanciotti, J. T. Roehrig, V. Deubel, J. Smith, M. Parker, K. Steele, B. Crise, K. E. Volpe, M. B. Crabtree, J. H. Scherret, R. A. Hall, J. S. MacKenzie, C. B. Cropp, B. Panigrahy, E. Ostlund, B. Schmitt, M. Malkinson, C. Banet, J. Weissman, N. Komar, H. M. Savage, W. Stone, T. McNamara, e D. J. Gubler. Origin of the west nile virus responsible for an outbreak of encephalitis in the northeastern united states. **Science**, 286(5448):2333–2337, 1999.
- R.S. Lanciotti, D.J. Gubler, e D.W. Trent. Molecular evolution and phylogeny of dengue-4 viruses. **Journal of General Virology**, 78:2279–2286, 1997.

- J. A. Lewis, G. J. Chang, R. S. Lanciotti, R. M. Kinney, L. W. Mayer, e D. W. Trent. Phylogenetic relationships of dengue-2 viruses. **Virology**, 197(1):216–224, 1993.
- W. B. Messer, D. J. Gubler, E. Harris, K. Sivananthan, e A. M. de Silva. Emergence and global spread of a dengue serotype 3, subtype III virus. **Emerging Infectious Diseases**, 9(7):800–809, 2003.
- M. Miagostovich. **Informes em Saúde Pública 2: Simpósio sobre Dengue**. Capítulo Características dos vírus dengue e de suas replicações, páginas 3–11, Universidade De São Paulo, 1999.
- R. J. P. Neto, D. M. Lima, S. O. de Paula, C. M. Lima, I. M. Rocco, e B. A. L. Fonseca. Molecular epidemiology of type 1 and 2 dengue viruses in Brazil from 1988 to 2001. **Brazilian Journal Of Medical And Biological Research**, 38(6):843–852, 2005.
- R. Nielsen e Z. Yang. Likelihood models for detecting positively selected amino acid sites and applications to the hiv-1 envelope gene. **Genetics**, 148:929–936, 1998.
- L. R. Petersen e J. T. Roehrig. West nile virus: A reemerging global pathogen. **Emerging Infectious Diseases**, 7(4):611–614, 2001.
- O. G. Pybus e A. Rambaut. Evolutionary analysis of the dynamics of viral infectious disease. **Nature Reviews Genetics**, 10(8):540–550, 2009.
- R. Rico-Hesse. Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. **Virology**, 174(2):479–493, 1990.
- R. RicoHesse, L. M. Harrison, R. A. Salas, D. Tovar, A. Nisalak, C. Ramos, J. Boshell, M. T. R. deMesa, R. M. R. Nogueira, e A. T. daRosa. Origins of dengue type 2 viruses associated with increased pathogenicity in the americas. **Virology**, 230(2):244–251, 1997.

- F. Ronquist e J. P. Huelsenbeck. MrBayes3: Bayesian phylogenetic inference under mixed models. **Bioinformatics**, 19:1572–1574, 2003.
- N.A. Rosenberg e M. Nordborg. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. **Nature Reviews – Genetics**, 3:380–390, 2002.
- A. Rudnick. Studies of ecology of dengue in Malaysia. **Bulletin Of The World Health Organization**, 35(1):78–, 1966.
- F. Tajima. Evolutionary relationship of dna sequences in finite populations. **Genetics**, 105(2):437–460, 1983.
- S. Tavaré. Line of descent and genealogical processes and their applications in population genetics models. **Theoretical Population Biology**, 26(2):119–164, 1984.
- J. D. Thompson, D. G. Higgins, e T. J. Gibson. Clustal-w improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, 22(22):4673–4680, 1994.
- S. S. Twiddy, E. C. Holmes, e A. Rambaut. Inferring the rate and time-scale of dengue virus evolution. **Molecular Biology And Evolution**, 20(1):122–129, 2003a.
- S. S. Twiddy, C. H. Woelk, e E. C. Holmes. Phylogenetic evidence for adaptive evolution of dengue viruses in nature. **Journal Of General Virology**, 83:1679–1689, 2002.
- S. Susanna Twiddy, Oliver G. Pybus, e Edward C. Holmes. Comparative population dynamics of mosquito-borne flaviviruses. **Infection, Genetics and Evolution**, 3(2):87–95, 2003b. ISSN 1567-1348.

- N. Vasilakis, E. J. Shell, E. B. Fokam, P. W. Mason, K. A. Hanley, D. M. Estes, e S. C. Weaver. Potential of ancestral sylvatic dengue-2 viruses to re-emerge. **Virology**, 358(2):402–412, 2007.
- E. Y. Wang, H. L. Ni, R. L. Xu, A. D. T. Barrett, S. J. Watowich, D. J. Gubler, e S. C. Weaver. Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. **Journal Of Virology**, 74(7):3227–3234, 2000.
- S. C. Weaver e A. D. T. Barrett. Transmission cycles, host range, evolution and emergence of arboviral disease. **Nature Reviews Microbiology**, 2(10):789–801, 2004.
- S. C. Weaver e N. Vasilakis. Molecular evolution of dengue viruses: Contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. **Infection, Genetics and Evolution**, 9(4):523 – 540, 2009. ISSN 1567-1348.
- World Health Organization WHO. Dengue, dengue haemorrhagic fever and dengue shock syndrome in the context of the integrated management of childhood illness. Relatório técnico, Department of Child and Adolescent Health and Development, Geneva, 2005.
- A. Wilder-Smith e D. J. Gubler. Geographic expansion of dengue: The impact of international travel. **Medical Clinics Of North America**, 92(6):1377–, 2008.
- V. Wittke, T. E. Robb, H. M. Thu, A. Nisalak, S. Nimmannitya, S. Kalayanrooj, D. W. Vaughn, T. P. Endy, E. C. Holmes, e J. G. Aaskov. Extinction and rapid emergence of strains of dengue 3 virus during an interepidemic period. **Virology**, 301(1):148–156, 2002.
- Z. Yang. **Computational Molecular Evolution**. OSEE, 2006.
- Z. Yang, N. Goldman, e A. Friday. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. **Molecular Biology and Evolution**, 11:316–324, 1994.

- Z. Yang, R. Nielsen, N. Goldman, e A.K. Pedersen. Codon-substitution models for heterogeneous selection pressure at amino acid sites. **Genetics**, 155:431–449, 2000.
- Z. Yang, W.S.W. Wong, e R. Nielsen. Bayes empirical bayes inference of amino acid sites under positive selection. **Molecular Biology and Evolution**, 22: 1107–1118, 2005.
- P. M. A. Zanotto, E. A Gould, G. F. Gao, P. H. Harvey, e E. C. Holmes. Population dynamics of flavivirus revealed by molecular phylogenies. **Proceedings of the National Academy of Sciences**, 93:548–553, 1996.
- C. L. Zhang, M. P. Mammen, P. Chinnawirotpisan, C. Klungthong, P. Rodpradit, P. Monkongdee, S. Nimmannitya, S. Kalayanarooj, e E. C. Holmes. Clade replacements in dengue virus serotypes 1 and 3 are associated with changing serotype prevalence. **Journal Of Virology**, 79(24):15123–15130, 2005.
- Yang Zhou e Edward C. Holmes. Bayesian estimates of the evolutionary rate and age of Hepatitis B virus. **Journal of Molecular Evolution**, 65(2):197–205, AUG 2007. ISSN 0022-2844.

Apêndice A

Banco de Dados: DENV Database

Banco de Dados pode ser definido como uma coleção compartilhada de dados logicamente relacionados (Elmasri e Navathe, 2000). Arquivos contendo informações, a organização lógica e estruturada dessas informações e ferramentas para acesso a elas constituem as principais características de um banco de dados. As atuais técnicas de extração e sequenciamento de DNA geraram uma enorme quantidade de informações tornando imprescindível o uso de banco de dados biológicos de usos genéricos ou com informações específicas como, por exemplo, sequências virais.

Entre os bancos de sequências virais, um dos mais conhecidos é o Los Alamos (<http://www.hiv.lanl.gov/content/index>) na qual armazena informações referentes ao vírus HIV, influenza (H1N1) e hepatite HCV. Entretanto para vírus da dengue, nenhuma disponível atualmente na web é capaz de fornecer dados atualizados e com ampla distribuição geográfica. Sendo assim, a opção de dados para vírus da dengue disponível encontra-se no GenBank.

A construção do banco de dados relacional Denv-Database consistiu em três etapas de modelagem: modelo conceitual, lógico e físico. Os detalhes de cada um dos modelos são descritos a seguir.

A.1 Modelo Conceitual

Etapa inicial onde são delimitadas especificações no banco para representar a realidade. Seu entendimento é fácil para o usuário comum. O modelo

conceitual está fundamentado nos conceitos de entidade, relacionamento e atributos representados através do diagrama Entidade Relacionamento (ER) que utiliza elementos de retângulos, losangos e segmentos de reta. Os retângulos representam classes de objetos, os losangos são os relacionamentos e segmentos de retas definem atributos. Para construção do modelo conceitual foi utilizada a ferramenta brModelo (Candido, 2005).

Retângulos: círculos preenchidos,

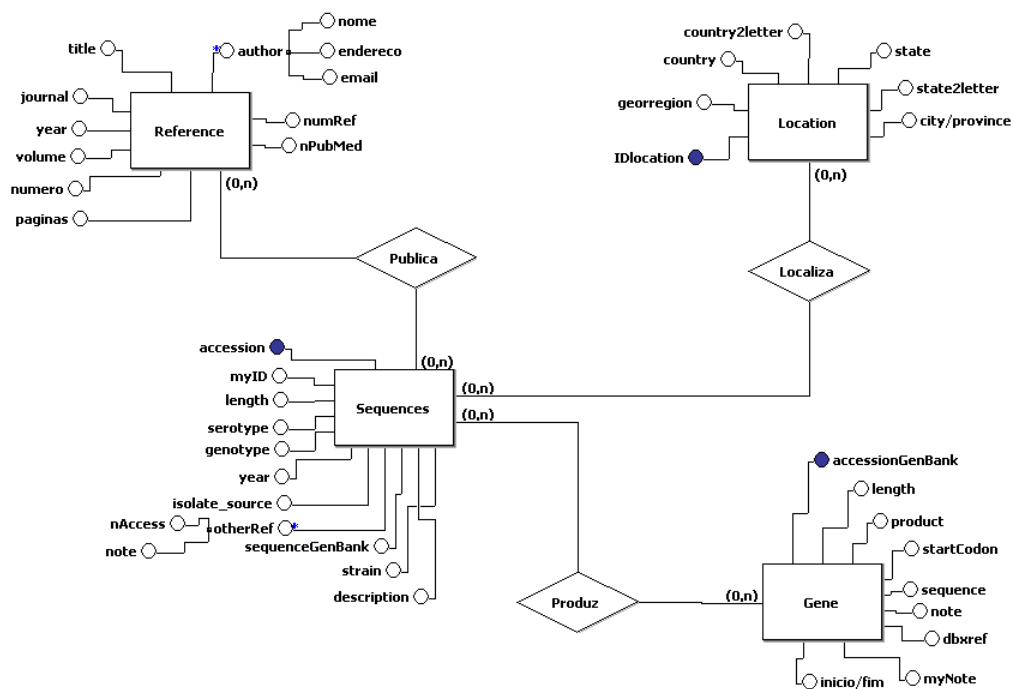


Figura A.1: Diagrama Entidade Relacionamento.

A.2 Modelo Lógico

O modelo lógico está sujeito às restrições do SGBD utilizado e pode sofrer eventuais reestruturações de forma a garantir a eficiência das consultas. Nesta etapa o Denv-Database foi projetado com o uso da ferramenta Designer que permitiu exportar o modelo diretamente para uma base de dados SGBD MySQL.

O resultado final consistiu de quatro tabelas principais: Referência,

Sequência, Genes e Locais amostrados. A tabela de Referência contém dados como autor, ano, título, jornal e ano de publicação. A tabela referente a Locais fornece informações da localidade onde foi extraída a amostra como país, região e continente. A tabela Sequência contém dados de sorotipo, ano de coleta, genotipo, tamanho, número de acesso, etc. Por fim, a tabela Gene apresenta informações derivadas de genomas completos (Figura A.2).

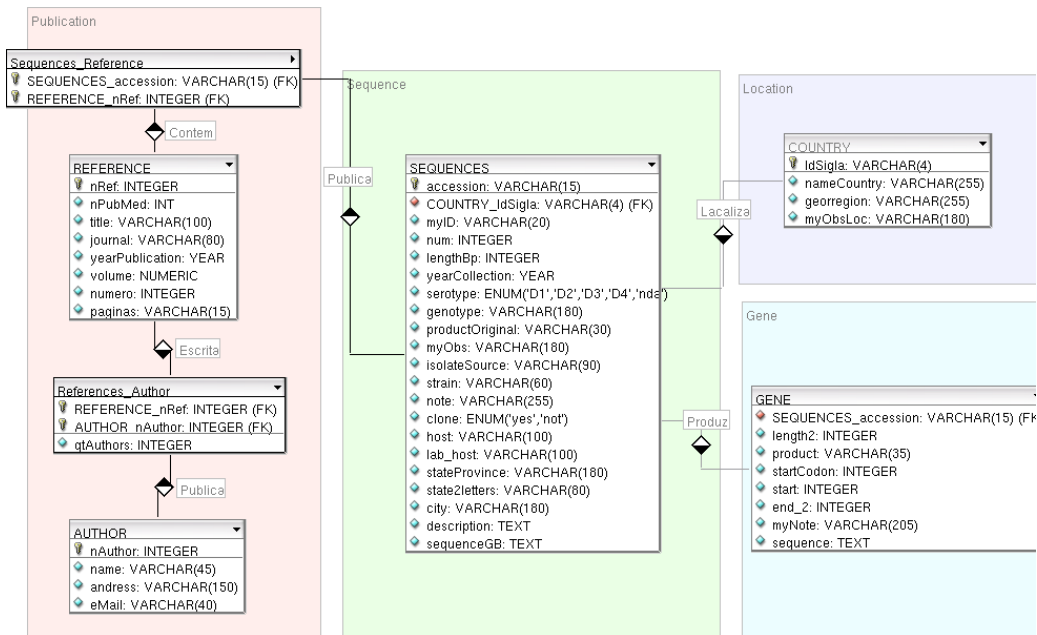


Figura A.2: Esquema do Modelo Lógico.

A.3 Modelo Físico

Esta etapa é dependente do SGBD a ser usado. O modelo físico define a organização e estruturas físicas de acesso aos dados (índices, métodos de acesso, etc.). Para o desenvolvimento do banco de dados, optou-se por trabalhar com o SGBD relacional MySQL devido à licença gratuita para uso acadêmico, leve e rápido (principalmente na plataforma Linux) e muito utilizado em aplicações web.

A.4 Dados e Estatísticas

Todas as informações das sequências que popularam a base de dados foram obtidas diretamente do GenBank e importados dentro da base local em MySQL usando scripts do módulo Bioperl. As sequências escolhidas tiveram mais que 100 nucleotídeos e são todas referentes ao vírus da dengue.

O total de dados no Denv-Database foi de 7089 nas quais 91.5

Quanto à região genômica, ENV apresentou 1597 sequências e mais 2103 associadas a registros do genoma completo. O total de 2056 amostras contém genomas completos de Denv muitos dos quais não constam separação de regiões codificantes CDS.

A.5 Consultas

Uma das principais vantagens de SGBD's está na facilidade de consultas e recuperação de informações. Primeiramente o acesso as informações foi feito através da ferramenta phpMyAdmin com objetivo de coletar informações da região do ENV que constituiu o conjunto de dados para análises evolutivas.

Posteriormente com o intuito de disponibilizar os dados para consulta na rede, foi feito um formulário na qual é possível obter informações referentes a subtipo, ano, localidade dentre outras (Figura A.3).

A.6 Portal Web

O portal web ainda está em fase de construção sendo assim, a interface gráfica e as operações sofrerão a inserção de novas funcionalidades, com o intuito de permitir consultas mais específicas e mais inteligentes (Figura A.4). O portal utiliza PHP v.5.3 para conectar a base de dados e gerar dinamicamente um ambiente amigável em HTML através do servidor Apache de acordo com o padrão de W3C para assegurar a interoperabilidade quando possível.

Sequence Search Interface

Accession	<input type="text"/>	Length	<input type="text"/>	pb a
Gene	<input type="text" value="Any"/>	Year	<input type="text"/>	a
Serotype	<input type="text" value="Any"/>	collection	<input type="text"/>	
LOCATION :		Genotype	<input type="text"/>	
		Georregion	<input type="text" value="Any"/>	
		Country	<input type="text"/>	
		State	<input type="text"/>	
		City/Province	<input type="text"/>	
OTHER :		Strain	<input type="checkbox"/>	Reference
				<input type="checkbox"/>
		Isolate Source	<input type="checkbox"/>	
		Host	<input type="checkbox"/>	Description
				<input type="checkbox"/>
		Note	<input type="checkbox"/>	
OUT PUT :		List	<input type="text" value="Any"/>	records per page.
		Submit	Reset	

Figura A.3: Formulário de Consulta.

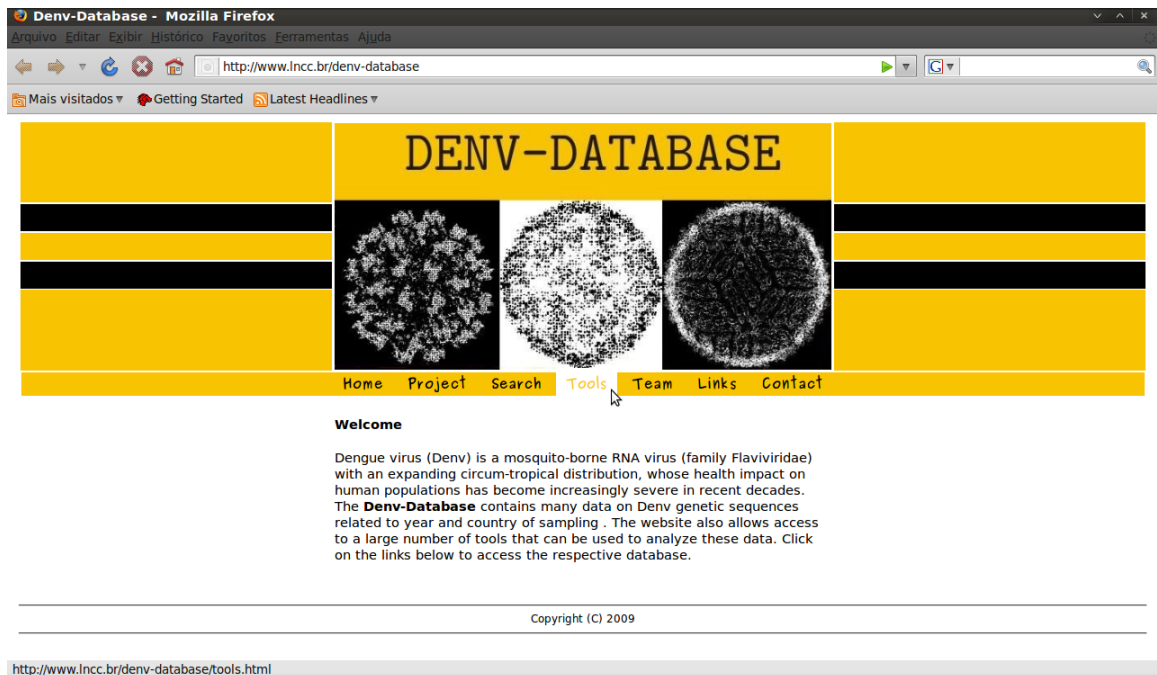


Figura A.4: Aplicação Web.

A.7 Ferramentas

Foi adicionado ao banco de dados Denv-Database a ferramenta de alinhamento BLAST (Basic Local Alignment Search Tool) que permite encontrar

similaridade entre uma sequência fornecida pelo usuário e as sequências disponíveis no banco de dados Altschul et al. (1990). Esta ferramenta auxilia na checagem do genótipo da sequência fornecida e verificação da origem da amostra.

Apêndice B

Lista das Siglas e Países Utilizadas nas Análises de Acordo com a ISO 3166-1

A ISO 3166-1 sugere código de duas letras para países e dependências. Localidades utilizadas no estudo estão listadas a seguir. Para demais localidades, consultar em: (<http://www.iso.org.iso>).

Sigla	Pais	Região	Sigla	Pais	Região
AI	Anguila	América Central	MQ	Martinique	América Central
AO	Angola	África	MS	Montserrat	América Central
AU	Australia	Oceania	MX	Mexico	América Central
AW	Aruba	América Central	MY	Malasia	Asia
BB	Barbados	América Central	MZ	Mozambique	África
BD	Bangladesh	Asia	NC	Nova Caledonia	Oceania
BF	Burquina Farso	África	NG	Nigeria	África
BO	Bolivia	América do Sul	NI	Nicaragua	América Central
BR	Brasil	América do Sul	NR	Nauru	Oceania
BS	Bahamas	América Central	PE	Peru	América do Sul
BZ	Belize	América Central	PF	Pol. Francesa	Oceania
CI	Costa do Marfim	África	PG	Papua Nova Guine	Oceania
CK	Ilhas Cook	Oceania	PH	Filipinas	Sudeste Asiático
CN	China	Asia	PR	Porto Rico	América Central
CO	Colombia	América do Sul	PY	Paraguai	América do Sul
CR	Costa Rica	América Central	RE	Reunion	África
DM	Dominica	América Central	SA	Arabia Saudita	Oriente Médio
DO	Rep. Dominicana	América Central	SB	Ilhas Salomao	Oceania
EC	Equador	América do Sul	SC	Seychelles	África
FJ	Fiji	Oceania	SG	Singapura	Sudeste Asiático
GD	Grenada	América Central	SN	Senegal	África
GH	Ghana	África	SO	Somalia	África
GY	Guiana	América do Sul	SR	Suriname	América do Sul
HN	Honduras	América Central	SV	El Salvador	América Central
ID	Indonesia	Sudeste Asiático	TH	Tailandia	Sudeste Asiático
IN	India	Asia	TL	Timor Leste	Sudeste Asiático
JM	Jamaica	América Central	TO	Tonga	Oceania
JP	Japao	Asia	TT	Trinidad e Tobago	América Central
KH	Cambodia	Sudeste Asiático	TW	Taiwan	Asia
KM	Comodoros	África	UG	Uganda	África
KN	S.Cristovão e Nevis	América Central	US	Estados Unidos	América do Norte
KR	Coreia do Sul	Asia	VE	Venezuela	América do Sul
LC	Santa Lucia	América Central	WS	Samoa	Oceania
LK	Sri Lanka	Asia	VN	Vietname	Sudeste Asiático
MG	Madagascar	África	VI	Ilhas Virgens Am.	América Central
MM	Myanmar	Asia			

Apêndice C

Sequências do GenBank

D1-1945-US-AF425619	D1-1994-SA-AM746220	D1-2007-NI-FJ898433	D1-1997-TH-AY732391	D1-2004-TH-EU48386
D1-1944-US-X76219	D1-2004-KR-EF654104	D1-2008-NI-FJ547088	D1-1998-TH-AY732472	D1-2000-TH-AY732407
D1-1958-TH-D10513	D1-1977-GD-AF425618	D1-2005-NI-FJ873814	D1-1998-JP-AB111064	D1-1944-US-EU848545
D1-1963-TH-AF425629	D1-1977-JM-AF425621	D1-1987-PR-FJ478458	D1-2001-MM-AY606062	
D1-1972-MY-AF231721	D1-1978-TT-AF425631	D1-1994-PR-FJ410179	D1-2001-CN-EF508200	
D1-1974-NR-U88536	D1-1986-TT-AF425639	D1-1999-PY-AB111065	D1-2001-TH-FJ850068	
D1-1974-PH-AF425627	D1-1980-MX-AF425623	D1-1998-PR-EU482567	D1-2003-TH-EU448393	
D1-2003-PH-EU448406	D1-1983-MX-AF425624	D1-1991-PE-AF425626	D1-2007-MY-EU448395	
D1-2006-KR-EF654110	D1-1982-BR-AF425613	D1-1994-US-AY780642	D1-2004-CN-AY871812	
D1-2007-PH-EU448405	D1-1988-AO-AF425610	D1-2000-BR-FJ850071	D1-2007-TH-EU448394	
D1-2002-PH-AY422778	D1-1993-CR-AY153755	D1-2002-BR-FJ850075	D1-2000-KH-FJ639669	
D1-2001-PH-AY422777	D1-1996-PR-FJ478457	D1-2005-BR-FJ850084	D1-2001-KH-FJ639670	
D1-2003-ID-EU448404	D1-1985-CO-AF425616	D1-2007-BR-FJ850090	D1-2004-KH-FJ639682	
D1-2006-ID-EU448402	D1-1986-PR-FJ562106	D1-2003-BR-FJ850077	D1-2002-SG-EU069600	
D1-1977-JM-D00501	D1-1995-PR-FJ547086	D1-2004-BR-FJ850081	D1-2003-SG-EU069598	
D1-1985-AW-D00505	D1-1992-PR-FJ547087	D1-2006-BR-FJ850087	D1-2004-SG-EU069593	
D1-1983-AU-AF425611	D1-1993-PR-FJ562105	D1-2006-IN-EU448413	D1-2005-SG-EU448399	
D1-2000-TL-EF440432	D1-1996-CO-AF425617	D1-1986-TH-AY732432	D1-2006-CN-EF508207	
D1-2002-CN-EF508201	D1-1994-VE-AF425637	D1-1987-TH-AY732425	D1-2006-MY-EU448400	
D1-2007-ID-EU448409	D1-1995-VE-AF425638	D1-1987-TW-AF425628	D1-2007-CN-FJ158609	
D1-2002-ID-AB232666	D1-1998-VE-FJ639741	D1-1985-TH-AY732426	D1-2005-KR-EF654105	
D1-2002-JP-AB111073	D1-1997-VE-AF425634	D1-1982-TH-AY732397	D1-2005-KH-FJ639685	
D1-2005-ID-EU448407	D1-1999-VE-FJ639743	D1-1992-TH-AY732395	D1-2006-KH-FJ744702	
D1-1995-CN-D0855297	D1-2008-VE-FJ850104	D1-1988-TH-AY732380	D1-2003-KH-FJ850069	
D1-2004-MY-EU448410	D1-2006-PR-EU482591	D1-1991-TH-AY732382	D1-2002-KH-FJ639674	
D1-2001-PF-FJ898448	D1-2006-VE-FJ639815	D1-1993-TH-AY732465	D1-2006-VN-EU249490	
D1-2004-RE-EU282328	D1-2004-VE-FJ639794	D1-1994-TH-AY732468	D1-2007-KH-FJ639696	
D1-2006-MG-EU448412	D1-2005-VE-FJ639808	D1-1999-MM-AY618878	D1-2001-JP-AB111066	
D1-2004-SC-DQ285561	D1-2007-VE-FJ873809	D1-2000-MM-AY618877	D1-2003-VN-EU482476	
D1-1968-NG-AF425625	D1-2008-BR-FJ850093	D1-1981-TH-AY732383	D1-2004-VN-FJ882569	
D1-1985-CI-AF425620	D1-1997-BR-AF425614	D1-2005-SA-AM746214	D1-2005-VN-FJ882570	
D1-1976-MM-AF425615	D1-2001-BR-FJ850073	D1-2006-SA-AM746213	D1-2008-VN-FJ410213	
D1-1980-TH-AY732447	D1-2004-NI-EU596501	D1-2004-SA-AM746216	D1-2007-VN-EU448388	
D1-1983-TH-AY732379	D1-2006-NI-FJ024479	D1-1995-TH-AY732387	D1-1996-TH-AY732455	
D1-1993-KM-DQ285562	D1-2006-SV-EU448414	D1-2002-TH-AY732403	D1-1999-TH-AY732470	

D2-2004-NI-FJ898436 D2-2001-BR-FJ850074 D2-1983-JM-DQ364484 D2-1974-TH-AJ487271 D2-1976-ID-L10044-6 D2-1990-LK-L10049
D2-2006-NI-EU482444 D2-2002-BR-FJ850076 D2-1988-TH-DQ181846 D2-1982-TH-DQ181819 D2-1991-SG-AF410368 D2-1989-LK-L10055
D2-2005-NI-EU482597 D2-2004-BR-FJ850082 D2-1991-TH-AF195040 D2-1983-TH-DQ181817 D2-1998-TH-AF410377 D2-1996-SR-FJ882602
D2-1999-SV-DQ364515 D2-2005-BR-FJ850085 D2-1997-VN-AF410349 D2-1986-TH-DQ181831 D2-1988-MX-DQ341200 D2-1994-LK-AF410372
D2-2000-CR-DQ364514 D2-2001-PY-EU045311 D2-1998-VN-AF410345 D2-1994-TH-DQ181850 D2-1994-JP-AB111448 D2-1992-SA-AF410378
D2-1999-MX-DQ341201 D2-2005-PY-EU045312 D2-2002-KH-FJ639698 D2-1984-TH-DQ181834 D2-2000-JP-AB111450 D2-1994-SA-AM746227
D2-2007-NI-EU482601 D2-1992-MQ-DQ364519 D2-2004-KH-EU448419 D2-1985-TH-DQ181832 D2-1993-AU-AY037116 D2-1993-UG-AF410375
D2-2008-NI-FJ810418 D2-1991-PR-DQ364531 D2-2001-KH-FJ639697 D2-1987-TH-DQ181829 D2-2000-TL-EF440433 D2-1991-IN-AF410371
D2-2000-NI-FJ898478 D2-1993-PR-EU482589 D2-1998-MM-DQ518638 D2-1990-TH-DQ181843 D2-2004-ID-DQ518634 D2-1994-IN-AF410374
D2-2001-NI-FJ744704 D2-1989-PR-EU482562 D2-2005-TW-DQ518640 D2-1993-TH-DQ181886 D2-2000-SG-EU069583 D2-2004-IN-EU448422
D2-2002-MX-FJ898438 D2-1990-PR-EU482586 D2-2005-VN-DQ518639 D2-1995-TH-DQ181881 D2-2001-SG-EU069582 D2-2004-LK-AB194883
D2-2008-MX-FJ898439 D2-1988-PR-DQ364482 D2-2002-VN-FM210227 D2-1996-TH-FJ906958 D2-2005-SG-EU069581 D2-1996-JP-AB111449
D2-2002-BZ-FJ898461 D2-1987-PR-EU482576 D2-1999-VN-FM210217 D2-1999-TH-DQ181875 D2-2002-MY-FM986657 D2-2004-BD-EU448424
D2-1991-VE-AF398107 D2-1992-PR-EU482594 D2-2003-VN-DQ518641 D2-2000-SV-DQ364560 D2-1999-MY-FM986656 D2-2005-BD-EU448423
D2-1999-VE-AF398108 D2-1994-PR-DQ364478 D2-2001-VN-FM210238 D2-2001-TH-DQ181871 D2-2002-SG-EU069572 D2-2006-IN-FJ898454
D2-2000-VE-AF398110 D2-1986-PR-DQ364546 D2-2004-VN-DQ518642 D2-2000-TH-DQ181894 D2-2003-SG-EU069573 D2-1992-AU-AF410370
D2-2003-VE-FJ639783 D2-1990-VI-FJ898450 D2-1989-TH-AF195041 D2-1997-TH-DQ181900 D2-2004-SG-EU069577 D2-2004-SA-AM746222
D2-2006-VE-FJ639822 D2-1997-TT-DQ364559 D2-1944-CO-EU854293 D2-2003-TH-FJ898452 D2-2003-MY-DQ518635 D2-2004-CN-AY871813
D2-2004-CO-FJ024477 D2-2000-EC-DQ364511 D2-1944-CO-FJ390389 D2-2008-VN-FJ410215 D2-2006-ID-EU448431 D2-1969-PR-AF264054
D2-2004-VE-FJ850112 D2-1993-CO-DQ364512 D2-2008-PG-FJ906959 D2-2006-VN-FJ873811 D2-2004-MY-DQ518636 D2-1969-PR-L10046
D2-2005-VE-FJ639732 D2-1996-CO-AF163096 D2-1981-TW-EF540856 D2-2007-VN-EU448417 D2-2007-CN-FJ158608 D2-1974-TO-X54319
D2-2005-CO-EU854294 D2-1997-CO-DQ364497 D2-1983-PH-L10045 D2-2006-TH-EU448416 D2-2005-ID-DQ518637 D2-1977-PR-DQ917243
D2-2007-VE-EU482604 D2-2000-PR-EU687222 D2-1995-PH-AY512569 D2-2007-KH-EU448415 D2-2007-SG-EU448429 D2-1984-HN-FJ898449
D2-2008-VE-FJ850106 D2-2001-PR-EU687232 D2-1998-PH-AY512568 D2-2008-KH-FJ639718 D2-2001-JP-AB111454 D2-1995-PE-DQ917242
D2-2001-DO-AB122020 D2-2005-PR-EU482556 D2-1996-PH-AF295694 D2-2004-TH-DQ518646 D2-2005-PH-DQ518631 D2-1957-IN-L10043
D2-2005-VI-FJ898453 D2-2006-PR-EU482544 D2-2003-PH-EU448418 D2-2005-KH-FJ639711 D2-2001-TW-DQ518630 D2-1953-TT-L10053
D2-2007-BR-FJ850091 D2-2004-PR-EU529693 D2-1987-TW-L10052-6 D2-2003-KH-DQ518649 D2-1997-MY-FM986655 D2-1970-MY-AF231717
D2-2003-DO-FJ898451 D2-2002-PR-EU482723 D2-1994-PH-DQ518643 D2-1998-TW-DQ518645 D2-2001-PH-DQ518632 D2-2008-MY-FJ467493
D2-2007-PR-EU596491 D2-2003-PR-EU687199 D2-1958-TH-D10514 D2-1992-TH-DQ181859 D2-2004-PH-AB194884 D2-1980-CI-DQ917244
D2-1998-MQ-AF208496 D2-1995-PR-EU569706 D2-1964-TH-NC_00147 D2-2002-MM-DQ518651 D2-2007-PH-EU448426 D2-1980-SN-AF231718
D2-1998-PR-EU687212 D2-1997-PR-DQ364509 D2-1974-TH-DQ181816 D2-2004-MM-DQ518652 D2-2007-ID-EU448425 D2-1980-BF-DQ917247
D2-1999-PR-EU677145 D2-1996-PR-EU677141 D2-1977-TH-DQ181812 D2-1983-BF-L10042-6 D2-1997-CK-AF004020 D2-1970-SN-AF231720
D2-2001-KN-FJ898460 D2-1996-VE-EU687220 D2-1976-TH-DQ181815 D2-1984-SO-L10051-6 D2-1977-SC-L10047
D2-2003-BR-FJ850078 D2-1998-VE-FJ898465 D2-1981-TH-DQ181825 D2-2005-GH-EU005258 D2-1977-SC-L10048
D2-2006-BR-FJ850088 D2-2000-BR-FJ850072 D2-1976-TH-DQ181814 D2-1996-AU-AF004019 D2-1985-LK-L10050

D3-1956-PH-L11423 D3-1998-TW-DQ518662 D3-1995-WS-FJ898456 D3-2000-PR-FJ562107 D3-2002-PE-FJ898458
 D3-1956-PH-M25277 D3-1993-TH-AY676357 D3-1996-PH-DQ518668 D3-2002-PR-EU687196 D3-2004-PE-DQ177888
 D3-1973-JP-AB111085 D3-2001-TH-FJ810414 D3-1998-PH-DQ518669 D3-2001-PR-FJ547079 D3-2005-PE-DQ177897
 D3-1962-TH-L11440 D3-1996-VN-DQ518655 D3-1994-TW-DQ518667 D3-2005-PR-FJ182007 D3-2003-PE-DQ177891
 D3-1973-TH-L11620 D3-1998-TH-AY676388 D3-1997-PH-DQ518670 D3-2007-PR-EU596492 D3-2000-EC-DQ177898
 D3-1974-TH-AY676360 D3-1998-VN-DQ518654 D3-2005-TW-DQ518672 D3-2000-VE-FJ639746 D3-2000-BR-AY038605
 D3-1977-TH-AY676384 D3-2000-TH-AY676408 D3-2007-PH-EU448432 D3-2003-CO-FJ898443 D3-2003-BR-DQ118867
 D3-1983-TH-AY676358 D3-2003-TH-DQ518660 D3-2005-PH-DQ518673 D3-2001-VE-DQ367720 D3-2006-BR-FJ850089
 D3-1982-TH-AY676355 D3-1999-TH-AY676413 D3-2006-PH-EU448433 D3-2004-VE-FJ639801 D3-2007-BR-FJ850092
 D3-1980-TH-AY676359 D3-2004-KH-FJ639726 D3-1998-ID-DQ518676 D3-2002-VE-FJ639775 D3-2008-BR-FJ850094
 D3-1990-TH-AY676361 D3-2005-KH-FJ639727 D3-2003-ID-DQ518677 D3-2007-CO-FJ898445 D3-2005-BR-FJ850086
 D3-1985-TH-AY676368 D3-2006-KH-EU448440 D3-2005-ID-DQ518678 D3-2003-VE-DQ367721 D3-2006-PY-EU045323
 D3-1991-TH-AY676416 D3-2008-KH-FJ639715 D3-2006-MY-EU448439 D3-2006-VE-FJ639825 D3-1999-PR-EU529696
 D3-1987-TH-AY676362 D3-2004-VN-DQ518656 D3-2007-MY-EU448438 D3-2005-CO-FJ898444 D3-2006-PR-EU482555
 D3-1984-TH-AY676371 D3-2007-KH-FJ639712 D3-2006-ID-EU448435 D3-2008-VE-FJ639826 D3-2004-PR-FJ850056
 D3-1998-MM-EU448447 D3-2003-VN-DQ518657 D3-2007-ID-EU448434 D3-2006-MX-FJ898441 D3-2007-VE-FJ898474
 D3-2000-JP-AB111080 D3-2005-VN-DQ518658 D3-2005-SG-EU448436 D3-2007-MX-FJ898442 D3-2005-VE-EU854292
 D3-2001-BD-DQ518665 D3-2007-VN-FJ410229 D3-2000-TL-EF440434 D3-2003-MX-FJ898440 D3-2008-NI-FJ898476
 D3-2000-BD-AY656669 D3-2006-VN-EU448441 D3-2005-TL-DQ453968 D3-2002-PY-EU045316 D3-1963-PR-L11433-6
 D3-2006-BD-EU448446 D3-2008-VN-FJ461322 D3-1985-SR-FJ882574 D3-2004-BR-DQ118882 D3-1963-VE-AY146762
 D3-2007-BD-EU448445 D3-1973-ID-L11425 D3-1989-SR-FJ882572 D3-2000-MQ-AY099338 D3-1977-VE-AY146761
 D3-2005-MM-DQ518666 D3-1974-MY-L11429 D3-1985-LK-L11436 D3-1999-MQ-AY099337 D3-1965-PF-L11439
 D3-2007-MM-EU448443 D3-1978-ID-L11426 D3-1981-LK-L11431 D3-2004-PY-EU045322 D3-1977-PR-L11434
 D3-2006-MM-EU448444 D3-1985-ID-L11428 D3-1984-IN-L11424 D3-2001-LC-FJ898463
 D3-1992-TH-AY676386 D3-1991-ID-DQ518674 D3-1985-MZ-FJ882575 D3-2002-TT-FJ898459
 D3-1994-TH-AY676366 D3-1997-MY-FM986662 D3-1989-LK-L11437 D3-2001-MQ-AY099340
 D3-1995-TH-AY676373 D3-1999-TW-DQ518675 D3-1991-LK-L11438 D3-2002-GY-FJ898464
 D3-2002-TH-AY676383 D3-1983-PH-L11432-6 D3-1999-LK-DQ518679 D3-2002-BR-DQ118881
 D3-2000-KH-FJ639719 D3-1981-MY-L11427-6 D3-2000-LK-AY099336 D3-2003-PR-FJ390372
 D3-2001-KH-FJ639720 D3-1989-NC-EU182246 D3-1993-SR-FJ882573 D3-2003-BO-DQ177887
 D3-2003-KH-FJ639723 D3-1989-PF-EU182239 D3-2004-SA-AM746229 D3-2003-PY-EU045317
 D3-2002-KH-FJ639721 D3-1991-CK-FJ898455 D3-1997-SA-AM746228 D3-2001-BR-FJ898446
 D3-1996-TH-AY676365 D3-1990-PF-EU182240 D3-1986-WS-L11435 D3-2001-AI-FJ898462
 D3-1997-TH-DQ518661 D3-1992-FJ-L11422-6 D3-1994-NI-FJ882576 D3-2000-PE-DQ177903
 D3-1996-JP-AB111084 D3-1996-PF-EU182241 D3-1998-PR-FJ547070 D3-2001-PE-DQ177900

D4-1961-JP-AB1111090 D4-2003-TH-EU448457 D4-1998-PR-FJ882598
D4-1973-JP-AB1111091 D4-2005-TH-EU448456 D4-1994-MS-AY152369
D4-1956-PH-U18433 D4-2007-TH-EU448454 D4-1999-BB-AY152368
D4-1964-PH-U18434 D4-2005-KH-EU448455 D4-1998-BS-AY152364
D4-1984-PH-U18435 D4-1973-ID-U18428-5 D4-1998-VE-FJ639739
D4-2003-PH-EU448449 D4-1978-CN-EF436282 D4-1999-VE-FJ639745
D4-2005-PH-EU448448 D4-2001-SG-EU448464 D4-2000-VE-FJ639748
D4-1963-TH-U18440 D4-2002-MY-FM986674 D4-1993-BB-AY152376
D4-1969-MY-AF231722 D4-2002-JP-AB111088 D4-1994-PR-FJ226067
D4-1978-LK-U18437 D4-2007-SB-EU448462 D4-1995-PR-FJ810417
D4-1976-TH-AY618949 D4-2004-ID-EU448461 D4-1994-SR-AY152372
D4-1977-TH-AY618950 D4-2006-ID-EU448460 D4-2007-VE-EU854299
D4-1980-TH-AY618952 D4-2007-ID-EU448459 D4-2008-VE-FJ882592
D4-1978-TH-U18441 D4-2004-PH-EU448458 D4-2001-VE-FJ639773
D4-1981-TH-AY618953 D4-2000-TL-EF440435 D4-1997-MY-FM986673
D4-1985-TH-AY618961 D4-1976-ID-U18429 D4-1999-MY-FM986671
D4-1983-TH-AY618956 D4-1977-ID-U18430 D4-1997-TH-AY618979
D4-1984-TH-AY618958 D4-1981-JM-AY152389 D4-1973-MY-AF231724
D4-1982-TH-AY618955 D4-1982-SR-AY152385 D4-1975-MY-AF231723
D4-1986-TH-AY618962 D4-1983-JM-AY152384 D4-1975-MY-AF231725
D4-1996-IN-AB111086 D4-1984-NC-U18432
D4-1995-TH-AY618974 D4-1983-SV-U18426
D4-1996-TH-AY618977 D4-1981-DM-AF326573
D4-1993-TH-AY618969 D4-1982-BR-U18425
D4-1998-TH-AY618980 D4-1979-PF-U18438
D4-1999-JP-AB111087 D4-1984-MX-U18431
D4-1999-TH-AY618984 D4-1991-HN-AY152379
D4-1991-TH-AY618964 D4-1991-MX-AY152378
D4-2002-TH-AB111089 D4-1996-MX-DQ341217
D4-2003-KH-EU448453 D4-1994-SV-U18427
D4-2004-VN-EU448452 D4-1985-PF-U18439
D4-2002-VN-EU448451 D4-1997-CO-FJ024476
D4-2006-VN-EU448450 D4-1986-PR-EU854295
D4-2000-TH-AY618935 D4-1999-PR-FJ882599
D4-2001-TH-AY618944 D4-1996-PR-FJ850058

Apêndice D

Bases de dados Especializadas em Virus da dengue

VÍRUS DE RNA:

<http://bioinf.cs.auckland.ac.nz/rnavirusdb/>

DENGUE DT DATABASE:

Disponível em: <http://www.bioinformatics.org/dengueDTDB/Pages/main.htm>

BROAD INSTITUTE:

<http://www.broadinstitute.org/annotation/viral/Dengue>

DENGUE DATABASE:

http://www.denguedb.org/dengue_info.asp

Apêndice E

Modelos Substituições de Nucleotídeos

Todos os métodos discutidos a seguir assumem as seguintes premissas:

- (i): Todos os sítios de nucleotídeos mudam independentemente;
- (ii): A taxa de substituição é constante através do tempo e em diferentes linhagens;
- (iii): A composição de bases está em equilíbrio;
- (iv): A probabilidade condicional de substituição do nucleotídeo é a mesma para os sítios e não mudam através tempo.

Jukes Cantor 69

Foi o primeiro modelo proposto para substituição de nucleotídeos, sendo considerado o mais simples. Este modelo envolve um único parâmetro a ser estimado, assim, ele também é conhecido como modelo um parâmetro (*single-parameter model*) e assume que as quatro bases têm frequências iguais e todas as substituições são igualmente prováveis. Neste modelo, a distância de duas sequências é dada por:

$$d = -\frac{3}{4} \ln\left(1 - \frac{4}{3}p\right) \quad (\text{E.1})$$

onde p é a proporção de nucleotídeos que são diferentes em duas sequências e \ln é uma função log natural. A matriz que representa a probabilidade de substituição e o vetor de composição de bases é dada por:

$$P_t = \begin{bmatrix} . & \alpha & \alpha & \alpha \\ \alpha & . & \alpha & \alpha \\ \alpha & \alpha & . & \alpha \\ \alpha & \alpha & \alpha & . \end{bmatrix} f = \left[\frac{1}{4} \frac{1}{4} \frac{1}{4} \frac{1}{4} \right] \quad (\text{E.2})$$

onde α é a probabilidade de substituição que é a mesma para todas as possíveis substituições.

Os demais modelos estão descritos no apêndice D.

Kimura 80

Neste modelo é incorporado diferenças observadas entre as taxas de transições e transversões. Com base nestas informações, foi proposto o modelo Kimura dois parâmetros que incorpora a taxa de transição por sítio (α) diferente da taxa de transversão (β), dada a taxa de substituição total por sítio de $(\alpha + 2\beta)$. A taxa transição/transversão (α/β) pode ser representada pela letra kappa (κ).

A distância é dada por:

$$d = \frac{1}{2} \ln[1 - 2P - Q] + \frac{1}{4} \ln[1/(1 - 2Q)] \quad (\text{E.3})$$

onde P e Q são as diferenças proporcionais entre duas sequências devido a transição e transversão respectivamente. A matriz de probabilidades de transição e a frequência de bases é dada por:

$$P_t = \begin{bmatrix} . & \beta & \alpha & \beta \\ \beta & . & \beta & \alpha \\ \alpha & \beta & . & \beta \\ \beta & \alpha & \beta & . \end{bmatrix} f = \left[\frac{1}{4} \frac{1}{4} \frac{1}{4} \frac{1}{4} \right] \quad (\text{E.4})$$

Felsenstein 81

Neste modelo, alguns nucleotídeos podem ter uma frequência maior que outros e é de se esperar que algumas substituições sejam mais frequentes que outras. A composição de bases pode variar entre genes e entre o mesmo gene

em diferentes espécies. Assim, o modelo Felsenstein permite que a frequência dos quatro nucleotídeos sejam diferentes. A matrix de substituição é dado por:

$$P_t = \begin{bmatrix} . & \pi_C\alpha & \pi_G\alpha & \pi_T\alpha \\ \pi_A\alpha & . & \pi_G\alpha & \pi_T\alpha \\ \pi_A\alpha & \pi_C\alpha & . & \pi_T\alpha \\ \pi_A\alpha & \pi_C\alpha & \pi_T\alpha & . \end{bmatrix} f = [\pi_A\pi_C\pi_G\pi_T] \quad (\text{E.5})$$

onde π_i é a frequência da i -ésima base calculados sob a sequência que são comparados.

HKY 85

Este modelo foi proposto por Hasegawa, Kishino e Yano em 1985. Essencialmente ele é uma junção dos modelos K2P e F81 por permitir a ocorrência em diferentes taxas da transição e tranversão e também a variação na frequência de bases. Sua matrix de substituição e frequência de bases é dada por:

$$P_t = \begin{bmatrix} . & \pi_C\beta & \pi_G\alpha & \pi_T\beta \\ \pi_A\beta & . & \pi_G\beta & \pi_T\alpha \\ \pi_A\alpha & \pi_C\beta & . & \pi_T\beta \\ \pi_A\beta & \pi_C\alpha & \pi_T\beta & . \end{bmatrix} f = [\pi_A\pi_C\pi_G\pi_T] \quad (\text{E.6})$$

Comparação de modelos

Apêndice F

Breve Descrição dos Programas

Utilizados

F.1 BEAST

A arquitetura global do pacote BEAST (*Bayesian Evolutionary Analysis by Sampling Trees*) é um pipeline. Beast é um programa geral de inferência bayesiana por MCMC dos parâmetros de modelos evolutivos quando as árvores são coalescentes. Inclui uma variedade de modelos dentre os quais, modelos de substituição de nucleotídeos, tempo de divergência das amostras, relógio molecular relaxado e modelos populacionais. BEAST foi escrito como um executável Java capaz de rodar em qualquer sistema com Java 1.4 ou superior. A entrada de arquivo é no formato xml com a descrição dos dados para análise.

Disponível gratuitamente em: <http://beast-mcmc.googlecode.com>

Programas relacionados ao BEAST:

F.1.1 Beauti

Beauti (*Bayesian Evolutionary Analysis Utility*) é um software gráfico que permite a criação de arquivos para entrada no Beast (formato xml).

F.1.2 TreeAnnotator

Sintetiza a informação de uma amostra de árvores produzidas no BEAST em uma árvore de alvo único orientado para enraizada medidas filogenéticas de tempo usando modelos de relógio relaxado ou molecular estrito. Destina-se tanto como um método de reconstrução de filogenias e como uma estrutura para testar a hipótese evolutiva sem condicionamento em um BEAST única árvore topologia usa MCMC a média de espaço de árvore, de modo que cada árvore é ponderada proporcional à sua probabilidade posterior. Aplicação nas áreas de filogenia de espécie para datação molecular genética de populações baseada em coalescência e medidas envolvendo populações (DNA ancestral ou data das sequências virais no conjunto de dados).

F.2 TRACER

Software de ferramentas gráficas para análises de saídas MCMC. Lê arquivos de saída gerados pelo BEAST, MrBayes e BAli-Phy. Também escrito em Java.

Disponível em: <http://www.evolve.zoo.ox.ac.uk/software.html?id=tracer>

F.3 PAML

Constitui um pacote de programas feito por Ziheng Yang para a análise de máxima verossimilhança de nucleótidos ou proteínas sequências, incluindo os métodos baseados em códons que levem em conta tanto os aminoácidos quanto nucleótideos. Os programas podem estimar comprimentos dos ramos em uma árvore filogenética e os parâmetros do modelo evolutivo como a razão da taxa transição/transversão (κ), o parâmetro de gama (α) para as taxas de substituição variável entre os sítios, os parâmetros de taxa substituições não sinônimas/sinônimas (ω). O pacote também pode testar modelos evolutivos, calcular as taxas de substituição em locais particulares, reconstruir nucleotídeos ancestrais ou sequências de aminoácidos e simular a evolução da sequência da proteína. O pacote está disponível como arquivos executáveis do Windows e de

código fonte C para Unix e sistemas MacOS X.

Disponível em: <http://abacus.gene.ucl.ac.uk/software/paml.html>

F.3.1 CODEML

É um conjunto de programas inseridos no pacote PAML com objetivo de realizar análises filogenéticas de DNA ou proteína utilizando método de máxima verossimilhança, (ML). Capaz de estimar taxas não sinônimas e sinônimas por sítio ou linhagem.

F.4 ClustalW

Programa de alinhamento automático de múltiplas sequências (Thompson et al., 1994). Realiza alinhamento cruzado de pares usando o algoritmo de Needleman e Wunsch 1970 para calcular os pares de distâncias, na qual usa para reconstruir uma árvore NJ. Então ele utiliza a árvore NJ como guia para o alinhamento progressivo das sequências. Existem duas variedades de Clustal: ClustalW com interface em linha de comando e ClustalX com interface gráfica. Roda na maioria das plataformas e o download podem ser obtidos através do link: <ftp://ftp.ebi.ac.uk>.

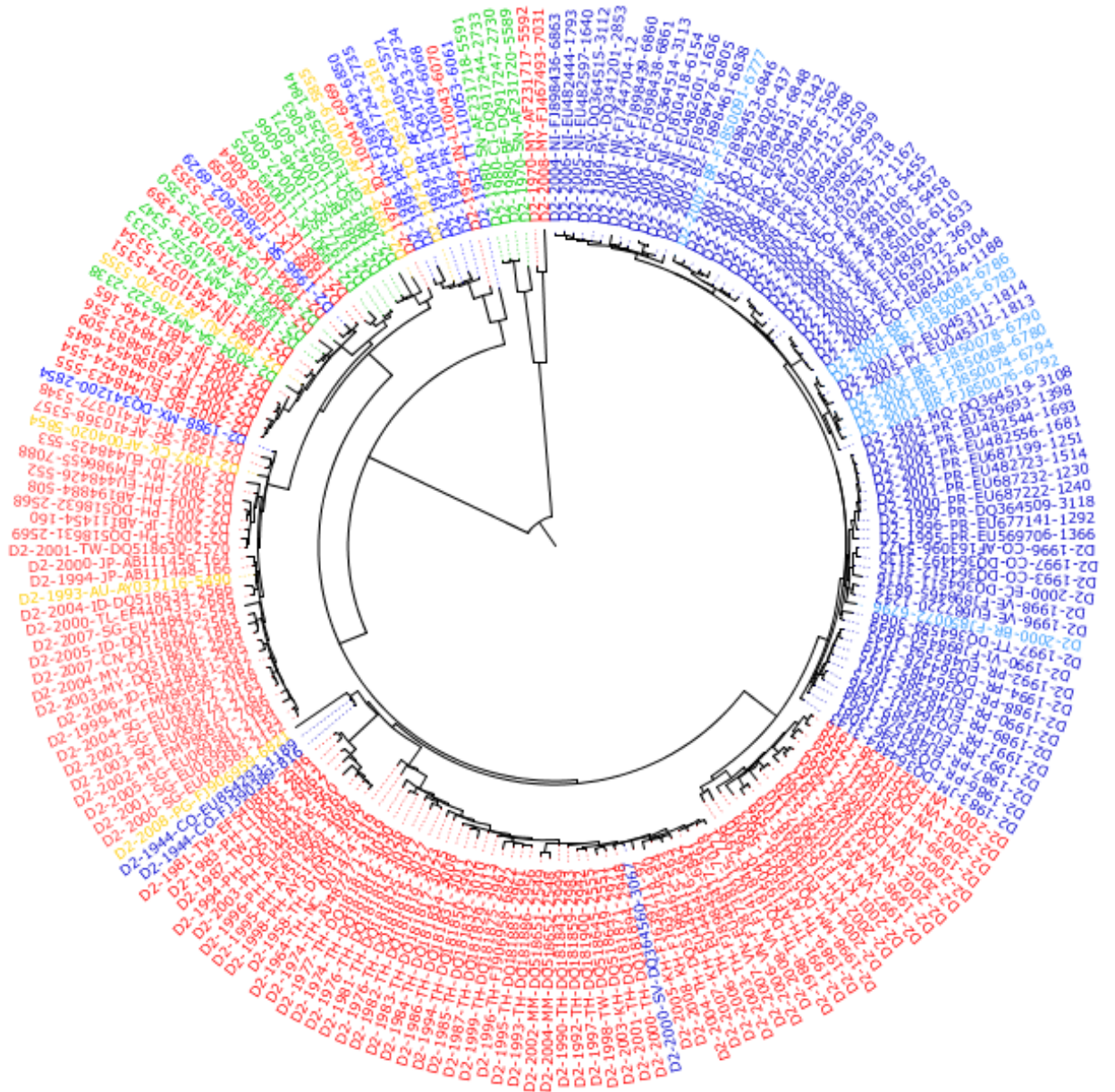
F.5 MEGA

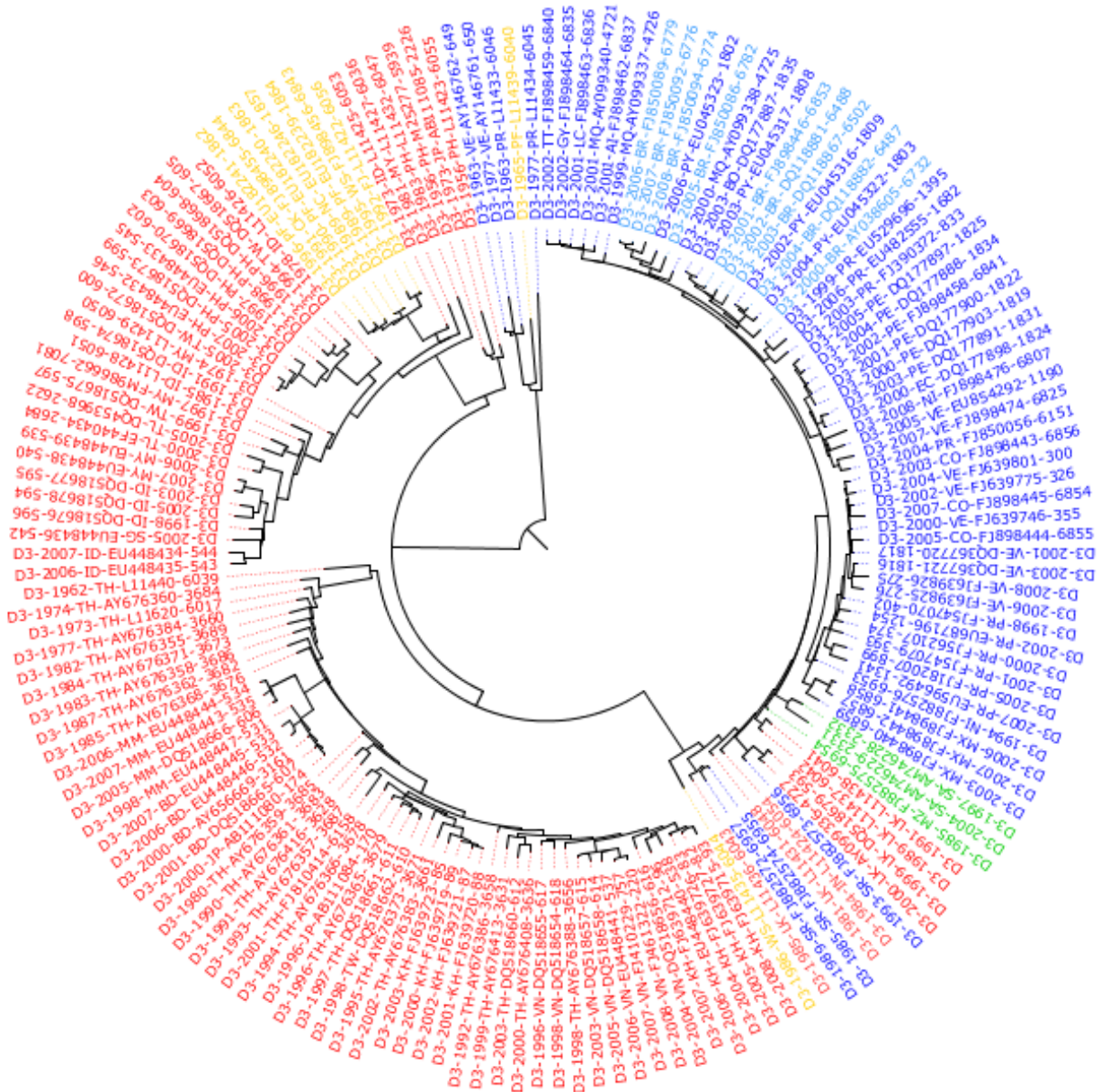
MEGA *Molecular Evolutionary Genome Analysis* é uma plataforma com ferramentas integradas capazes de realizar alinhamento de sequências de forma automática e/ou manual, edição de sequências, inferência de árvores filogenéticas, mineração bases de dados da web, estimativa de taxas da evolução molecular e testes de hipóteses evolutivas.

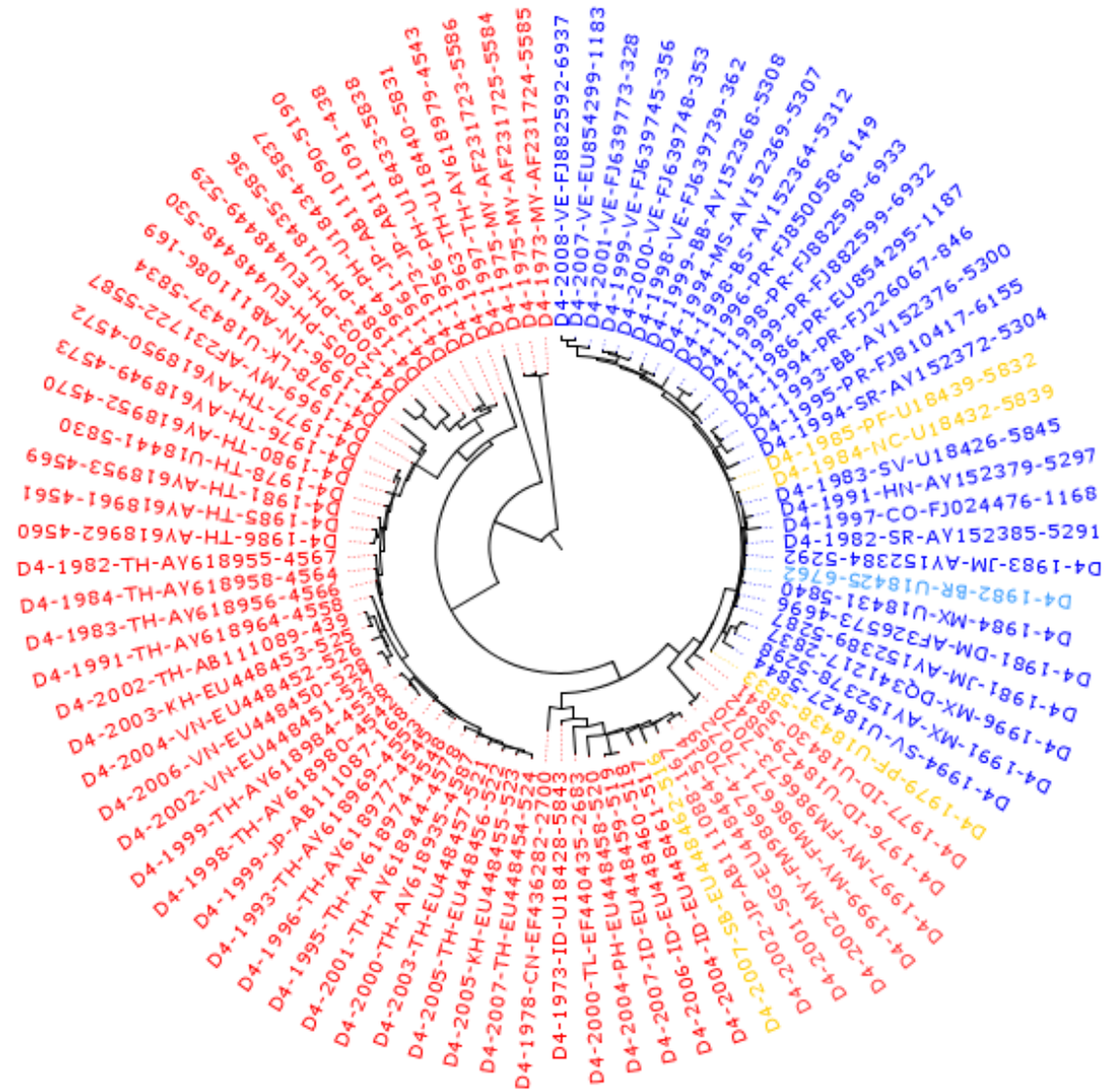
Disponível em: www.megasoftware.net

Apêndice G

Árvores Genealógicas Circulares







Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)