

**SELEÇÃO NATURAL NO VÍRUS DA HEPATITE C: INFLUÊNCIA DO SISTEMA  
IMUNE NA RESPOSTA AO TRATAMENTO**

**ARTUR TRANCOSO LOPO DE QUEIRÓZ**

TESE APRESENTADA  
AO  
PROGRAMA INTERUNIDADES EM BIOINFORMÁTICA  
DA  
UNIVERSIDADE DE SÃO PAULO  
PARA  
OBTENÇÃO DO GRAU DE DOUTOR  
EM  
CIÊNCIAS

Área de Concentração: **Bioinformática**

Orientador: **Prof. Dr. Sergio Russo Matioli**

Coorientadora: **Dra. Isabel Maria Vicente Guedes de Carvalho-Mello**

**São Paulo**

**2010**

**ARTUR TRANCOSO LOPO DE QUEIRÓZ**

**SELEÇÃO NATURAL NO VÍRUS DA HEPATITE C: INFLUÊNCIA DO SISTEMA  
IMUNE NA RESPOSTA AO TRATAMENTO**

Tese apresentada ao Programa Interunidades de Pós-  
graduação em Bioinformática da Universidade de São  
Paulo para obtenção do título de Doutor em Bioinformática  
Orientação: Prof. Dr. Sergio Russo Matioli  
Coorientação: Dra. Isabel Maria Vicente Guedes de  
Carvalho-Mello

**São Paulo**  
**2010**

**Dados Internacionais de Catalogação na Publicação (CIP).**

© reprodução autorizada pelo autor

Queiroz, Artur Trancoso Lopo de.

Seleção natural no vírus da hepatite C: A influência do sistema imune na resposta ao tratamento/ por Artur Trancoso Lopo de Queiroz. – 2009.

Tese(doutorado) – Universidade de São Paulo, Programa Interunidades de Pós-Graduação em Bioinformática, 2009.

Orientador: Prof. Dr. Sergio Russo Matioli.

Coorientadora: Dra. Isabel Maria Vicente Guedes de Carvalho-Mello.

Descritores: 1. HCV-1. 2. Epítomos. 3. Evolução. 4. Pressão seletiva positiva.

Dedicat6ria

**Dedico este trabalho à meu amado filho,  
Minha mãe e Minha avó,  
e aos meus avós**

**Agradecimientos**

À Deus por iluminar o caminho até aqui.

Ao meu amado filho por seu amor e carinho que sempre me deu e por ser a minha maior motivação, orgulho e alegria em todo o mundo. Obrigado por existir.

À minha Mãe e Minha avó, por terem me dado todo o apoio e força para continuar, pelo carinho e amor que sempre me fizeram seguir em frente e pelos exemplos que me foram dados.

Aos meus falecidos avôs e avó, sempre presentes em minha vida até o momento que se foram e que sempre farão falta no meu coração.

À minha grande amiga Flora, por ter me dado o apoio e a oportunidade de vir para São Paulo, um voo que parecia ser impossível, mas conseguimos!

À minha amiga e coorientadora Isabel, que me acolheu e me ajudou em momentos muito difíceis e que sempre acreditou em mim. Sem seu apoio essa tese não seria possível.

Ao meu orientador Sergio, que me aceitou e me ajudou nessa longa caminhada até aqui.

Aos meus amigos Papito (Lucas) e Velho (Heitor), por me acolherem quando eu cheguei aqui na selva de pedra. Sem a ajuda de vocês não teria conseguido.

À minha namorada Vanessa, pelo seu carinho, amor, apoio e atenção que me deu e por sempre estar lá quando eu precisei.

Ao meu amigo Vinícius, pela sincera amizade, apoio e grande ajuda nos trabalhos e nas disciplinas. Ele sabe bem o que a gente passou.

Ao meu amigo Thibério, pela sua grande ajuda, apoio e várias risadas compartilhadas

À nossa querida Patricia, protetora de todos os pós graduandos da Bioinformática, que sempre esteve disposta e sorridente quando solicitamos ajuda e nas horas difíceis e de desespero que precisamos desabafar.

Ao meu amigo Alexandre Paxchoal, pelo apoio e sua prestatividade nos momentos mais complicados da minha pós-graduação.

Ao meu amigo Dézão, por sempre me falar “relaxe não é tão difícil assim” quando eu estava prestes a entrar em desespero!

À Sibs “ai meu, ...meu” por sua sincera amizade e carinho que sempre mostrou ao ter nossas viagens paralelas.

À Priscila e Ricardo, por terem me ajudado imensamente quando eu estava a procura de apartamento gentilmente me acolhendo.

Aos meus amigos e colegas de apartamento Vânia e Helbert, foi muito divertido dividir o apartamento com vocês, mas sem a cãndida...

Aos meus amigos dos Peixes (Keila, Carol, Mari, Filipe, Mister Ringo, Marcinha, Ari...) pelos momentos divertidos enquanto esperava o programa correr.

Aos meus amigos das abelhas (Alayne, Elaine, Goianão (Paulo), Ex-Goianão(Leandro), Pato branco 1 (Alisson), Pato Branco 2(Gustavo), Flavio, Suzy) pelos momentos divertidos enquanto esperava o programa correr 2.

Ao pessoal do IMT (Laura, Cris (roda roda vira), Cris (RedBikini), Bixcoito, Jaila...) pelo apoio e amizade.

Aos nossos colaboradores de São José do Rio Preto, pela ajuda nos trabalhos realizados.

À Capes, pelo apoio financeiro.

E a todos aqueles que direta e indiretamente contribuíram para que este trabalho fosse realizado.

Epigrafe

*Veni, Vidi, Vici*

(Júlio César)

ou não...

(Caetano Veloso)

## **ABSTRACT**

The reason for low rates of sustained viral response (SVR) in HCV patients remains unknown. Several studies suggest that viral load is closely associated to viral and host factors, such as age, sex, body weight, transaminase levels and HCV genotype. Moreover, a strong CD8<sup>+</sup> T-cell immunity in acute resolving hepatitis C is matched by strong and sustained CD4<sup>+</sup> T-cell proliferation to multiple recombinant structural and non-structural viral proteins is responsible for the control of viraemia in HCV infection.

Here, we investigate the differences in CD8 epitopes frequencies from the Los Alamos database between groups of patients that showed distinct response to pegylated alpha interferon with ribavirin therapy, and test for evidence of natural selection on virus in the treatment failure group, using five maximum likelihood evolutionary models.

Our results indicated no evidence of positive selection by using pairwise models, however we identify evidence of positive selection in the non-responder group by applying maximum likelihood models. The epitope mapping showed that the epitope VLSDFKTWL was associated with effective therapy outcome, with statistical support.

Those results suggest that the immune response increases during the treatment period, allowing the virus clearance.

We also developed a tool for automatic mapping of epitopes of HCV that tests for evidence of natural selection by applying maximum likelihood models.

**Keywords:** C Hepatitis, Therapy, Epitopes, Positive selection.

## RESUMO

As taxas de resposta viral ao tratamento com Interferon- $\alpha$  associado com Ribavirina ainda não estão bem definidas. Muitos estudos associam vários fatores virais e do hospedeiro, tais como a idade, sexo, peso, etnia, nível de enzimas hepáticas, estágio de fibrose, genótipo do HCV com os níveis de RNA viral. Outros estudos associam as células CD8<sup>+</sup> específicas para epítomos do HCV com o controle da viremia. O objetivo desse trabalho foi verificar se há aumento na pressão seletiva contra o vírus nos pacientes em tratamento com Interferon- $\alpha$  associado à Ribavirina respondedores em comparação com os pacientes não respondedores e associar esse aumento com a ação antiviral do tratamento e/ou com o aumento da resposta imune devido à ação imunomoduladora do Interferon- $\alpha$ , utilizando-se modelos de máxima verossimilhança, de cálculo da razão dN/dS e realizando o mapeamento dos epítomos das proteínas NS5A.

Nossos resultados demonstram que usando modelos par a par não foi detectado evidência de seleção positiva, entretanto utilizando modelos de máxima verossimilhança nós observamos evidência no grupo que não responde a terapia. O mapeamento dos epítomos revelou que o epítomo VLSDFKTWL estava associado com a resposta efetiva do tratamento com suporte estatístico.

Estes resultados indicam que a resposta imunológica aumenta durante o período de tratamento, auxiliando na eliminação do vírus.

Além das análises, foi desenvolvida uma ferramenta de mapeamento automático dos epítomos do HCV e análise de seleção natural aplicando modelos de máxima verossimilhança.

**Descritores:** Hepatite C, Tratamento, Epítomos, Pressão Seletiva.

## SUMÁRIO

Lista de Abreviações

Lista de Tabelas

Lista de Figuras

1 – INTRODUÇÃO	1
1.1 - O vírus da Hepatite C	3
1.2 - Genoma e Proteínas Virais	4
1.3 - Características Filogenéticas	6
1.4 - Filodinâmica e Filogeografia do HCV	7
1.5 - Tratamento contra o HCV	9
1.6 - Resposta Imune contra o HCV	10
1.7 - Métodos para Análise de Evolução	12
1.8 - Ferramentas de Análise	14
2 – JUSTIFICATIVA	19
3 – OBJETIVOS	21
4 – MATERIAL E MÉTODOS	23
4.1 Amostragem	23
4.1.1 Pacientes infectados por HCV genótipo 1	23
4.1.2 Pacientes infectados por HCV genótipo 3	24
4.1.3 Pacientes transplantados	24
4.2 - Análise das sequências dos Pacientes	25
4.2.1 Pacientes infectados por HCV genótipo 1	25
4.2.2 Pacientes infectados por HCV genótipo 3	27
4.2.3 Pacientes transplantados Região NS3	27
4.3 - Ferramenta de Análise REEvolution	28
4.3.1 Fluxograma da ferramenta REEvolution	30
5 – RESULTADOS	33
5.1 Pacientes infectados por HCV de genótipo 1	33
5.2 Pacientes infectados por HCV de genótipo 3	36
5.3 Pacientes Transplantados Região NS3	37
5.4 Ferramenta de Análise	38
6 – DISCUSSÃO	41
7 – CONSIDERAÇÕES FINAIS	49
8 – REFERÊNCIAS	51
9 – ANEXOS	61

## **Lista de Abreviações:**

HCV – Vírus da Hepatite C

ISDR - Região determinante da sensibilidade ao Interferon.

PKR - Quinase de proteína ativada por RNA.

RT – Grupo dos pacientes respondedores ao tratamento.

MHC – Complexo principal de histocompatibilidade.

INF – Interferon.

NR – Grupo dos pacientes não respondedores ao tratamento.

NRFT – Grupo dos pacientes não respondedores de final de tratamento

PCR – Reação em cadeia da polimerase.

dS – Taxa de mutações sinônimas.

dN - Taxa de mutações não sinônimas.

dN/dS - Razão de mutações não sinônimas por mutações sinônimas.

BEB - Bayes empirical bayes.

UTR – Região não traduzida.

ORF – Quadro aberto de leitura.

RNA – Ácido ribonucleico.

HLA – Antígeno leucocitário humano.

PEG-INF- $\alpha$ -2b – interferon alfa ligado a PEG.

TCR – Receptor de célula T.

Ln – Logarítmico natural.

## **Lista de Tabelas:**

Frequência dos epítomos nas sequências da região NS5A do HCV dos grupos de pacientes (genótipo 1 a/b).	34
Resultados da análise de pressão seletora entre os grupos de pacientes.	35
Valores da razão dN/dS, dN e dS para cada epítomo nos grupos respondedor e não respondedor.	36
Mutações observadas nas regiões dos epítomos com significância estatística entre os grupos de pacientes.	37
Frequência dos epítomos nas sequencias da região NS3 dos pacientes transplantados infectados pelos genótipos 1 e 3.	38

## **Lista de Figuras:**

Figura 1 – Partícula viral do Vírus da Hepatite C.	4
Figura 2 – genoma e proteínas do Vírus da Hepatite C.	5
Figura 3 – Árvore filogenética de do HCV gerada a partir de sequencias da região NS5B	7
Figura 4 – Gráficos da análise de desempenho da ferramenta	39
Figura 5 – Interface da ferramenta <i>REEvolution</i>	46
Figura 6 – Exemplo de resultados gerados pela ferramenta <i>REEvolution</i>	47

# Introdução

## 1 – INTRODUÇÃO

A palavra vírus, no latim, significa veneno ou toxina. Atualmente esta palavra caracteriza pequenos agentes infecciosos que possuem material genético próprio mas são incapazes de se replicar sem estar inserido em uma célula viva de um organismo (Koonin *et al.*, 2006). Os vírus têm tamanhos que variam entre 20 e 300 nm de diâmetro e possuem representantes que infectam desde bactérias até animais e plantas.

A maioria dos vírus codifica enzimas próprias responsáveis pela replicação dos seus genomas de DNA ou RNA. Essas enzimas são polimerases que possuem uma taxa de erro na inserção de nucleotídeos maior que as polimerases de eucariotos, além de não possuírem atividade de correção. Esse fator, juntamente com a modificação de nucleotídeos causada por atividade de enzimas celulares do hospedeiro, a alta taxa de replicação viral e o processo de recombinação genética de alguns vírus, podem induzir diversos tipos de mutações além de inserções ou deleções no genoma viral (Domingo e Holland, 1997). Essas características observadas na maioria dos vírus explicam a evolução rápida apresentada por alguns tipos virais. Por exemplo, os vírus de RNA apresentam taxas de substituição nucleotídica que variam entre  $10^{-2}$  e  $10^{-5}$  substituições por sítio por ano (subs/sítio/ano), e a maioria possui taxas de  $10^{-3}$  subs/sítio/ano. No caso do vírus da Hepatite C, que possui um genoma de 9.400 nucleotídeos, é o equivalente à fixação de 9,4 substituições por genoma por ano (Jenkins *et al.*, 2002).

Essa informação é resultado de um aumento nos estudos sobre frequência de mutações realizados com Bacteriófagos, iniciados na década de 1940, que revelou que os vírus são bons sistemas para estudo de evolução e o conhecimento gerado sobre sua biologia molecular, replicação e ciclo celular permitem uma comparação com os processos evolutivos, fisiológicos e celulares de outros organismos. Além disso, o conhecimento das taxas de substituição e de fixação de mutações permite realizar análises comparativas de dados reais gerados em cultura com as hipóteses evolutivas e com o desenvolvimento de modelos de evolução mais realistas.

Entretanto, a rápida taxa de substituição durante o ciclo replicativo apresentada pelos vírus contribui significativamente para sua adaptabilidade. Essa característica dificulta a descoberta de novas regiões alvo para intervenções terapêuticas, pois frequentemente os vírus resistentes às vacinas e às drogas antivirais proliferam. Essa resistência é associada ao aparecimento de mutações de escape na região reconhecida pelo sistema imune ou na região com a qual o fármaco interage (Gerrish e Garcia-Lerma, 2003). A taxa de substituição

apresentada pelos vírus é muito elevada em comparação com aquelas dos primatas, que são estimadas em torno de  $2,2 \times 10^{-9}$  (Yi *et al.*, 2002). Os vírus de RNA possuem portanto uma taxa 100.000 vezes mais rápida, resultante de uma complexa interação de 4 fatores: A taxa de mutação basal do genoma viral, o tempo de geração do vírus, o tamanho populacional e a rápida fixação de mutações não sinônimas em relação às mutações sinônimas. Por exemplo, a rápida evolução do vírus da imunodeficiência humana (HIV) é reflexo de sua alta taxa de mutação aliada à presença frequente de pressão positiva seletiva e uma taxa de replicação elevada (Bonhoeffer *et al.*, 1995).

Apesar da grande importância do processo de mutação nos sistemas virais é relevante salientar a relação entre a taxa de mutação e a variação antigênica. A capacidade de gerar variantes de proteínas virais, reconhecidas pelo sistema imunológico, permite o escape do vírus mutantes da influência das moléculas e células efectoras. Este é um fator de grande importância na compreensão da dinâmica epidemiológica de uma virose e no desenvolvimento de vacinas eficientes. Assim, apesar dos vírus de RNA apresentarem semelhança nas taxas de substituição, existem grandes diferenças na capacidade de gerar diversidade antigênica. Por exemplo, o vírus do sarampo apresenta baixa diversidade antigênica apesar de sua taxa de substituição elevada. Essa característica particular permitiu o desenvolvimento de uma vacina estável. Entretanto, o vírus da influenza A demonstra grande diversidade antigênica, sendo necessárias atualizações contínuas nos antígenos presentes na vacina para que a mesma continue eficaz na proteção da população (Duffy *et al.*, 2008).

Esta grande disparidade entre a taxa de mutação e diversidade antigênica nos vírus de RNA é um dos grandes desafios no estudo de evolução viral. Por exemplo, o vírus da Hepatite C ainda não possui uma vacina eficaz e sua origem ainda não foi estabelecida.

### **1.1 – Vírus da Hepatite C (HCV)**

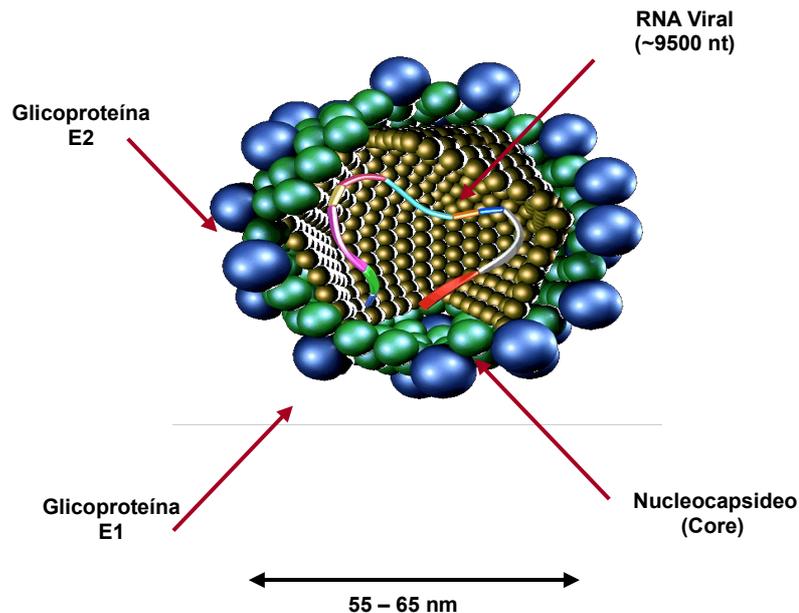
A hepatite crônica causada pelo vírus da hepatite C (HCV) constitui um grave problema de saúde pública mundial. Atualmente existem aproximadamente 170 milhões de pessoas no mundo (WHO, 1999) e em torno de 2 milhões de pessoas no Brasil infectadas por esse vírus (Ministério da Saúde, 2009). A infecção aguda é, na maioria dos casos, benigna e assintomática, o que dificulta o diagnóstico precoce. Entretanto, cerca de 70% desses casos evoluem para a infecção crônica, podendo progredir ao estágio de cirrose, doença hepática terminal e até para carcinoma hepatocelular (2,5%, segundo Bowen & Walker, 2005). É um vírus não citopático e hepatotrópico, atualmente classificado na família *Flaviviridae*.

Antigamente o HCV era classificado alternadamente nos gêneros *Flavivirus* e *Pestivirus*. Para o novo gênero foi proposto o nome *Hepacivirus* (Miller e Purcell, 1990; Shukla *et al.*, 1995). A partir dessa proposta, o nome *Hepacivirus* tem sido usado de forma disseminada.

A partícula viral mede de 55 a 65 nm de diâmetro (Kaito *et al.*, 1994) e é formada por um envelope lipídico, onde se encontram as glicoproteínas virais E1 e E2, um capsídeo protéico formado pela proteína viral *Core* e o genoma viral constituído por um ácido ribonucléico (RNA) linear, de fita simples, com polaridade positiva (Figura 1, Imbert *et al.*, 2004).

## 1.2 – Genoma e Proteínas Virais

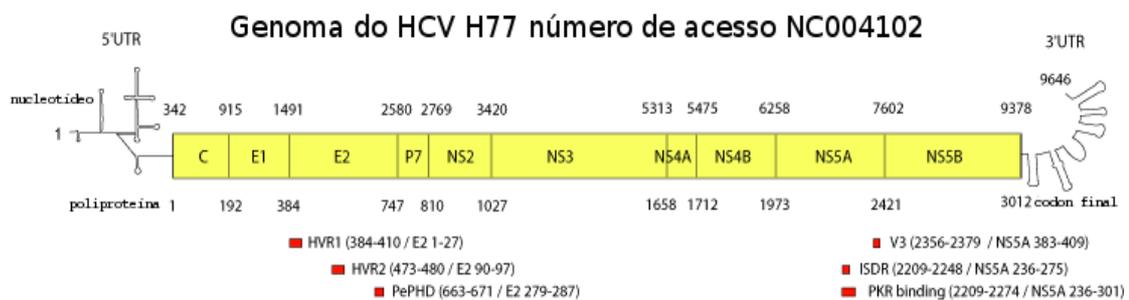
O RNA viral possui cerca de 9400 nucleotídeos com uma única fase aberta de leitura (ORF “open reading frame”), que abrange a maior parte do genoma e codifica uma poliproteína com pouco mais de 3000 resíduos de aminoácido, posteriormente clivada por proteases celulares e virais, que origina as proteínas do vírus.



**Figura 1** – Partícula viral do Vírus da Hepatite C. Fonte: (James, 2001), modificado.

Nas extremidades do genoma viral encontram-se duas regiões não codificantes denominadas 5'UTR e 3'UTR “Untranslated Region”, envolvidas na replicação do vírus. Na 5'UTR encontra-se um sítio de entrada interno para o ribossomo (IRES, “internal ribosomal entry site”) responsável pelo estabelecimento da tradução do RNA viral de forma independente do “CAP” 5', fundamental para a tradução dos RNAs mensageiros celulares.

As proteínas virais podem ser classificadas em 4 proteínas estruturais e 6 proteínas não estruturais. As proteínas estruturais são: A proteína do *core* ou nucleocapsídeo é a primeira da porção amino terminal da poliproteína. Ela resulta da clivagem da poliproteína nascente por proteases celulares formando a p21. Clivagens secundárias originam proteínas menores (p19 e p16) formando o capsídeo da partícula viral, as glicoproteínas do envelope, E1 (gp35) e E2 (gp70), que estão associadas com a entrada da partícula viral no hepatócito pela ligação com o receptor celular CD81 e são classificadas como proteínas transmembrânicas tipo I (Choo *et al.*, 1991; Lin *et al.*, 1994; Shukla *et al.*, 1995; Pileri *et al.*, 1998; Deleersnyder *et al.*, 1997; Bartosch *et al.*, 2003). Junto à proteína E2 há um polipeptídeo na sua extremidade carboxila, conhecido como p7, cuja origem ainda não está totalmente clara.



**Figura 2** – Genoma e proteínas do Vírus da Hepatite C. Modificado de: <http://hcv.lanl.gov/content/sequence/HCV/IMAGES/HCVgenome.png>.

As proteínas não estruturais são: NS2, NS3, NS4A e B, NS5A e B. A função das proteínas não estruturais ainda não está completamente elucidada. Entretanto, sabe-se que a proteína NS2 media sua própria clivagem em “cis” da proteína NS3, além de possuir também a função de metaloprotease. Clivagem em “cis” é aquela que ocorre apenas dentro da mesma molécula da poliproteína que catalisa a reação, enquanto clivagem em “trans” ocorre em outras moléculas da poliproteína. A proteína NS3 é uma proteína multifuncional, responsável pela clivagem de todas as proteínas não estruturais, com exceção da clivagem entre NS3 e NS4, além de estar relacionada com outros aspectos da infecção, pela associação com a quinase de proteína A ou PKA na transdução de sinais e no desenvolvimento de hepatocarcinoma. A NS4A funciona como cofator de NS3 e atua na fosforilação da NS5A. Contudo, a função da proteína NS4B ainda não está esclarecida. A proteína NS5A é a principal proteína não estrutural relacionada com a replicação do vírus e atua também se associando com a PKR, quinase de proteína induzida pelo Interferon, em que interfere na resposta do tratamento com Interferon alfa (INF- $\alpha$ ) (Bartenschlager, 2002; Katze *et al.*, 2002).

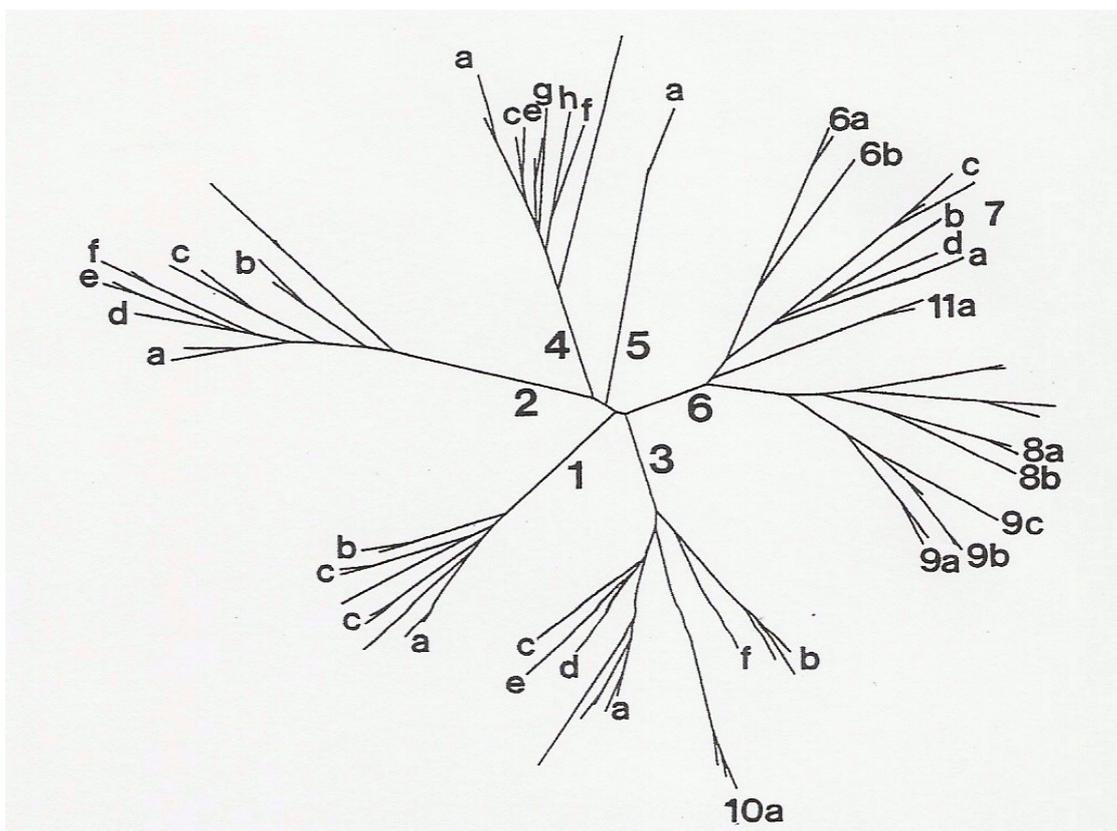
A proteína NS5B é uma polimerase de RNA dependente de RNA responsável pela replicação viral. Esta região possui um motivo GDD característico de uma replicase de RNA, presente em diferentes vírus de RNA (Hijikata *et al.*, 1993).

### 1.3 – Características Filogenéticas

O HCV apresenta ampla heterogeneidade genética. A comparação de sequências nucleotídicas, de diferentes isolados e de diferentes regiões geográficas, tem mostrado a existência de pelo menos seis grupos geneticamente diferentes (1-6) cuja variação de nucleotídeos do genoma completo nos diferentes grupos está entre 30 e 35% e a frequência de mutações é variável ao longo do genoma. Os grupos são constituídos por vários subtipos denominados por letras minúsculas (a, b, c, etc) com uma variabilidade nucleotídica de 20% a 25% entre eles (Simmonds *et al.*, 1994; Lau *et al.*, 1996). Essa grande divergência é reflexo da rápida taxa de substituição apresentada pelo vírus, sendo  $1,1 \times 10^{-3}$  substituições por sítio por ano, caracterizada recentemente no genótipo 1 (Magiorkinis *et al.*, 2009).

A distribuição geográfica dos genótipos varia bastante conforme a região, sendo que os genótipos 1a, 1b, 2a, 2b e 3a são cosmopolitas. Os genótipos 4, 5 e 6 parecem estar restritos a determinadas regiões. O genótipo 1 é o mais frequente mundialmente, sendo que a sua frequência varia entre 40% e 80%, nas regiões onde esse genótipo é presente. No Brasil, o padrão de distribuição dos genótipos segue o padrão mundial, sendo o genótipo 1 (65%) o mais frequente, seguido do 3 (30%), 2 (4%), outros (1%, Mellor *et al.*, 1995; Codes *et al.*, 2003; Campiotto *et al.*, 2005).

Em um indivíduo infectado, a população viral possui grande similaridade entre si, diferindo de 5 a 10%. O grupo de sequências altamente similares, derivadas de uma sequência dominante, é denominado de quasispécie. A variabilidade genética do HCV é atribuída, sobretudo, à baixa fidelidade da enzima polimerase de RNA dependente de RNA, que não possui atividade corretiva, permitindo, assim, a geração de variantes virais, que constituem as quasispécies (Martell *et al.*, 1992; Simmonds *et al.*, 2004). Outros fatores, como o escape da vigilância do sistema imunológico do hospedeiro e a alta taxa de replicação, demonstrada pelos estudos de cinética viral, podem também interferir na elevada diversidade das quasispécies encontradas em um indivíduo infectado (Zeuzem *et al.*, 1996).



**Figura 3** – Árvore filogenética radial do HCV gerada a partir de sequências da região NS5B (posição 7975-8196 em relação à H77). Fonte: Mizokami *et al.*, 1996.

#### 1.4 – Filodinâmica e Filogeografia do HCV

O conhecimento do genótipo infectante em cada paciente apresenta significativa relevância clínica. Os genótipos virais diferem com relação à sua agressividade e à taxa de resposta sustentada à terapia, ou seja, a negatização da carga viral mantida por seis meses após o término do tratamento (Lau *et al.*, 1996; Zhou *et al.* 1996).

A prevalência dos genótipos e subtipos virais é bem estabelecida, entretanto, suas origens ainda são desconhecidas. Padrões na diversidade viral sugerem que o HCV possui origem no Oeste africano ou no Sudeste asiático, entretanto nenhum vírus correlacionado foi encontrado em fontes não humanas, o que possibilitaria uma hipótese de uma zoonose (Simmonds *et al.*, 2005; Simmonds 2004, Simmonds 2001).

Estudos anteriores estimavam que o crescimento da epidemia do HCV se iniciou em torno de 1900, atingindo o ápice no final de 1980 (Smith *et al.*, 1997; Tanaka *et al.*, 2002; Nakano *et al.*, 2004). Entretanto esses resultados contrastam com as evidências epidemiológicas que sugerem que o crescimento da epidemia acompanha o crescimento de

procedimentos iatrogênicos parenterais (transfusões sanguíneas, uso de derivados de plasma, e injeções não descartáveis) e uso de drogas injetáveis, associadas com o uso compartilhado de seringas (Drucker *et al.*, 2001). Um estudo recente com o genótipo 1 demonstrou que a epidemia do HCV sofreu expansão durante os anos de 1940 e final de 1980 (Magiorokinins *et al.*, 2009), corroborando os dados epidemiológicos.

Recentemente foi demonstrado que a origem do genótipo 2 ocorreu no oeste africano, (Markov *et al.*, 2009), com origem na Guiné Bissau e Gâmbia, de onde se espalhou para o leste africano por Benim e Gana. O genótipo 2 é praticamente o único genótipo viral circulante na Guiné Bissau (Plamondon *et al.*, 2007). Além da migração interna no continente africano também foram encontradas evidências da migração desse genótipo para Martinica, no Caribe, durante os séculos XVII e XVIII, sugerindo que esse genótipo viral foi introduzido primariamente no Novo Mundo durante o período de tráfico de escravos. Entretanto, análises utilizando sequências de genótipo 2 não africanas demonstraram evidências de rotas mais recentes de migração viral no sul da Europa (especificamente no sul da França), o que seria uma alternativa à hipótese do tráfico de escravos (Markov *et al.*, 2009).

A origem do genótipo 1 ainda não foi estabelecida, porém foi descrito que sua disseminação ocorreu, provavelmente, para os países desenvolvidos (Austrália, Bélgica, França, Alemanha, Inglaterra, Grécia, Irlanda, Japão, Espanha, Suíça e Estados Unidos) e em seguida para os países em desenvolvimento (Argentina, Brasil, Camarões, China, Egito, Irã, Índia, Coreia, Rússia, Filipinas, Tailândia, Turquia, Uzbequistão e Vietnã), onde as epidemias locais foram estabelecidas posteriormente a partir de uma localização específica se espalhando devido à procedimentos iatrogênicos e uso de drogas injetáveis (Magiorokinins *et al.*, 2009; Stimson, 1996). Evidência similar já havia sido reportada como a migração do genótipo 1a proveniente dos Estados Unidos para América latina, associado ao aumento dos usuários de drogas injetáveis (Nakano *et al.*, 2004).

O perfil migratório do genótipo 3 ainda não foi estabelecido, entretanto a alta prevalência do genótipo 3a em jovens é associado com uso de drogas injetáveis (van Asten *et al.*, 2004; Schorter *et al.*, 2004). Essa característica sugere que o perfil migratório do genótipo 3 é semelhantes ao genótipo 1, porém nenhum estudo sobre a migração do genótipo 3 foi realizado.

## 1.5 – Tratamento contra o HCV

O tratamento da infecção crônica do HCV é baseada na combinação do interferon alfa ligado covalentemente à molécula de polietileno glicol (PEG-INF- $\alpha$ ) associado com ribavirina. Esse tratamento pode provocar reações adversas e cerca de 30-50% dos pacientes não apresentam resposta virológica sustentada (Thompson *et al.*, 2009), entretanto, ainda é o único tratamento minimamente eficaz contra a infecção. A ação do interferon é associada com o estabelecimento de um estado antiviral na célula infectada, porém sem a atuação direta na replicação viral. O INF afeta a tradução das proteínas virais ativando o gene da proteína quinase R. A fosforilação promovida por esta proteína bloqueia a síntese viral através da inibição do fator eucariótico de iniciação 2 (eIF2). A ribavirina é um análogo de guanosina, que atua diretamente na replicação do RNA viral. A incorporação de sua forma fosforilada, antagônica à citosina ou uracila, no RNA viral nascente resulta no bloqueio da polimerização do RNA. A ribavirina também atua na polimerase de RNA dependente de RNA provocando mutações no polímero sintetizado. O interferon e a ribavirina também apresentam atividade imunomoduladora. O interferon promove proliferação de células T de memória, inibe a apoptose das células T e aumenta a maturação das células dendríticas. Além disso, o interferon aumenta a expressão de moléculas do complexo principal de Histocompatibilidade (MHC) em todas as células e promove a modulação dos fenótipos das células CD4<sup>+</sup> para T<sub>h</sub>1, que é o fenótipo prevalente em respostas imunológicas específicas aos vírus e patógenos intracelulares. Além da atuação direta na replicação viral, a ribavirina também promove a modulação dos fenótipos das células CD4<sup>+</sup> para T<sub>h</sub>1 e no aumento na expressão de moléculas de MHC. A relevância do aumento observado na resposta imunológica durante o tratamento ainda não foi esclarecida (Feld e Hoofnagle, 2005).

Recentemente, com o desenvolvimento de sistemas de replicons do HCV, novos fármacos foram desenvolvidos para inibição mais direta da replicação viral. Atualmente existem inibidores que estão sendo avaliados através de testes clínicos, alguns deles em fases 2 e 3 (Thompson *et al.*, 2009). São os inibidores de protease (NS3/4A), SCH 503034 (boceprevir), VX-950 (telaprevir) e BILN 2061 (Tong *et al.*, 2008) e de polimerase (NS5B) NM283 (valopicitabine) e R1626 (Wohnsland *et al.*, 2007; Roberts *et al.*, 2008).

Estes agentes têm demonstrado um potente efeito antiviral que leva a uma redução significativa no RNA-HCV em um curto período de tratamento. Apesar do bom resultado que estas drogas vêm demonstrando em estudos preliminares, a rápida seleção de variantes virais demonstrando resistência fenotípica a estas drogas, associadas com mutações de escape, tem

sido observada em pacientes envolvidos em estudos clínicos ao se verificar elevação da carga viral durante o tratamento

## 1.6 – Resposta Imune contra o HCV

A palavra *immunitas* tem origem latina e refere-se à proteção contra os processos legais à qual os senadores romanos tinham durante o seu mandato. A palavra imunidade significa porém proteção contra doenças, especificamente contra doenças infecciosas. O sistema imunológico é composto de moléculas e células específicas cuja resposta é coordenada e coletiva.

Podemos dividir as resposta imunológica em dois tipos distintos: Resposta inata e Resposta adquirida.

A resposta inata, cujos mecanismos de defesa celulares e moleculares existem antes do estabelecimento de uma infecção. Seus componentes são as barreiras físicas (pele), as células fagocitárias (neutrófilos, monócitos e células dendríticas) e células NK (*Natural Killers*) ou matadoras naturais, proteínas envolvidas no sistema de complemento ou outros mediadores inflamatórios, e as citocinas, que regulam e coordenam as atividades celulares na imunidade natural.

A imunidade adquirida, cuja resposta é estimulada pela exposição à agentes infecciosos (de vírus e bactérias à helmintos e ácaros) sendo, dessa forma, uma resposta mais forte e específica do que a observada na resposta inata. Seus componentes são as células da linhagem linfóide, denominadas de linfócitos e seus produtos, como os anticorpos (Flajnik e du Pasquier, 2004). As moléculas não próprias, assim denominadas por não pertencerem ao repertório protéico do organismo, as quais a resposta imunológica específica é induzida são chamadas de antígenos (DNA, RNA, proteínas, carboidratos ou lipídeos).

A imunidade adquirida também pode ser dividida em dois tipos de resposta: Resposta celular e Resposta humoral.

A resposta celular, cujos componentes são as linhagens dos linfócitos T CD4<sup>+</sup>, que apresentam na sua superfície celular a molécula da família das imunoglobulinas CD4, e dos Linfócitos T CD8<sup>+</sup>, que apresentam na sua superfície a molécula CD8. Os linfócitos CD4<sup>+</sup>, também conhecidos como auxiliares, são os mecanismos efetores da imunidade celular adquirida responsável por coordenar os processos inflamatórios e intensidade de resposta imunológica, ativando outros tipos celulares por secreção de citocinas. Os linfócitos CD8<sup>+</sup>, também conhecidos como citotóxicos, são os responsáveis pela morte celular das células

infectadas por patógenos intracelulares. A morte celular ocorre através do reconhecimento do antígeno ligado ao complexo principal de histocompatibilidade do tipo I. Em seguida o linfócito libera as moléculas perforinas e caspases responsáveis pelo processo de apoptose.

A resposta humoral, cujos componentes são os linfócitos B e as moléculas da família das imunoglobulinas da classe Ig, chamadas de anticorpos. Essa é uma resposta imunológica específica contra patógenos extracelulares, na qual os anticorpos secretados pelos plasmócitos (linfócitos B ativos) reconhecem os antígenos presentes na superfície dos micro-organismos e promovem uma resposta inflamatória. As respostas celular e humoral são coordenadas, pois, para se tornar maduro, o linfócito B necessita de um sinal dos linfócitos T CD4<sup>+</sup> ativos (Silverstein, 2003).

O vírus da hepatite C é reconhecido pelos mecanismos da resposta imune inata, e induz uma rápida resposta por produção de interferons, o que gera um estado antiviral nas células hepáticas. Essa resposta é baseada em observações de picos de viremia em modelos com chimpanzé, onde se estima que apenas 10% dos hepatócitos abrigam a replicação do HCV (Bigger *et al.*, 2001; Wieland & Chisari, 2005). Esse vírus possui mecanismos que permitem-no escapar dessa resposta imune inata, como a protease NS3-4A, que bloqueia o reconhecimento de RNAs virais, e inibe a produção de INF- $\beta$  (Johnson & Gale, 2006). As proteínas NS5A e E2 inibem a enzima PKR induzida pelo interferon para a resposta antiviral (Taylor *et al.*, 1999; Gimenez-Barcons *et al.*, 2005). A ação das outras proteínas na evasão contra a resposta imune inata ainda não está clara.

Embora a resposta inata e a indução da produção de interferons sejam observadas após a infecção aguda, ela só é controlada por uma ativação extensa de células T específicas para o HCV (Thimme *et al.*, 2006; Racanelli *et al.*, 2003). Entretanto, a imunidade adquirida não consegue erradicar o HCV. A proteção contra a infecção recorrente do HCV depende das células CD4<sup>+</sup> e CD8<sup>+</sup> (Grakoui *et al.*, 2003; Shoukry *et al.*, 2003). Essas respostas são induzidas pelo reconhecimento do complexo MHC ligado à peptídeos antigênicos pelo receptor de célula T (*TCR*). Esses peptídeos antigênicos, oriundos de microorganismos, capazes de induzir resposta imunológica, são denominados epítomos (Terasaki, 2007). A resposta imune inata pode controlar a viremia na fase aguda. Contudo, mutações de escape têm sido relatadas em indivíduos crônicos e são relacionadas com a progressão para a infecção crônica (Erickson *et al.*, 2001). Alguns epítomos reconhecidos pelo HLA do hospedeiro passam a apresentar mutações de escape no momento em que é observado aumento na carga viral no paciente, o que sugere que as células CD8<sup>+</sup> promovam uma pressão seletiva nas proteínas virais, influenciando na replicação dos vírus.

Um exemplo são as mutações no epítipo imunodominante da proteína NS5B (epítipo com a sequência de resíduos de aminoácidos ARMILMTHF, no código de uma letra) que estão associadas com infecção persistente em indivíduos com HLA do tipo B27 (Neumann-Haefelin *et al.*, 2006). Essas mutações de escape mostram, aparentemente, que há um custo na aptidão darwiniana de replicação do vírus, pois, na ausência da pressão do sistema imune, a sequência do vírus se mantém estática ou retorna ao consenso do seu genótipo (Timm *et al.*, 2004; Cox *et al.*, 2005).

A resposta mediada por anticorpos não é associada ao controle da infecção aguda pelo HCV e o papel da resposta humoral ainda não foi esclarecido (Cooper *et al.*, 1999).

Estudos mostraram que o sucesso da terapia em pacientes crônicos está associado com o aumento de células T que reconhecem proteínas virais (Kamal *et al.*, 2002). A população de células T específicas para as proteínas do HCV aumenta durante o tratamento, independente do resultado deste (Barnes *et al.*, 2002), o que indica uma importante ação do sistema imune no controle da viremia na infecção pelo HCV. Entretanto, nem a terapia e sequer a resposta imune são 100% eficazes na erradicação do vírus.

### **1.7 – Métodos para Análise de Evolução.**

A mutação é um dos processos mais fundamentais na evolução viral no qual a seleção natural, deriva genética, fluxo gênico e recombinação atuam modificando a estrutura populacional. Na evolução Darwiniana, a seleção natural atua na variação genética de uma população existente de um determinado organismo maximizando a sua adaptabilidade “fitness” através da eliminação de variantes com mutações deletérias (seleção negativa) ou da proliferação de variantes com mutações favoráveis (seleção positiva; Orr, 2009).

Quando uma mutação é neutra, ou próxima da neutralidade, é esperado que a razão de ocorrência dessas mutações em um genoma seja semelhante à sua taxa de fixação no nível populacional (Li, 1993). Desvios observados nessa relação podem demonstrar influência de seleção natural. Dessa maneira podemos definir seleção natural positiva como o processo em que um novo mutante tem um valor adaptativo maior do que a média dos outros tipos pré existentes na população, o que ocasiona o aumento da frequência desse mutante nas gerações seguintes.

Cada aminoácido de uma proteína possui uma função biológica própria, logo, pode sofrer diferentes tipos e intensidades de forças seletivas. O número excessivo de mutações não sinônimas fixadas em uma população é indicativo de seleção natural positiva ocorrida

anteriormente. Da mesma maneira, o excesso de mutações sinônimas é considerado como resultante de ausência de seleção em um organismo que está adaptado ao ambiente, sendo indicativo de atuação pregressa de seleção natural negativa (Suzuki e Gojobori, 1999).

Consequentemente, Miyata e Yasunaga (1980) propuseram a “taxa de aceitação” para medir a evolução adaptativa no nível molecular, que permitiria entender os processos responsáveis pela adaptação na morfologia, comportamento e fisiologia, entre outros. Para entender o papel da seleção natural, a informação existente nas regiões codificadoras de proteínas oferecem uma vantagem sobre a informação contida em *introns* e sequências não codificantes, pois permitem identificar mutações sinônimas e não sinônimas. Mutações sinônimas são definidas como aquelas em que há uma mutação no DNA sem que haja mudanças na proteína, devido ao código genético ser degenerado. As mutações não sinônimas ocorrem quando a mutação no DNA resulta na modificação do aminoácido na proteína.

A razão da proporção de mutações não sinônimas/mutações sinônimas ( $\omega$  ou dN/dS) pode evidenciar se houve ou não pressão seletiva positiva no nível de proteína. Pressão seletiva negativa, ou purificadora, ocorre quando as mutações são deletérias e sua taxa de fixação é reduzida, o que significa que a proporção de fixação das mutações não sinônimas é menor que a fixação das sinônimas, ou seja, dN menor que dS levando a um  $\omega < 1$ . Da mesma maneira, evolução neutra é definida se a seleção não atua na aptidão (*fitness*), logo, as mutações não sinônimas serão fixadas na mesma taxa do que as mutações sinônimas e o dN será igual ao dS o que resulta em uma razão  $\omega = 1$ . Seleção positiva ocorre quando as mutações não sinônimas são favorecidas por seleção darwiniana, o que resulta em uma maior fixação de mutações não sinônimas em relação às mutações sinônimas, indicado por uma razão  $\omega > 1$ . Com a razão dN/dS é possível determinar se uma região está sob influência de um fator de seleção ou não. Para estimar a razão dN/dS, duas classes de métodos foram desenvolvidos para estimar dN e dS: Métodos de contagem Heurísticos e métodos de verossimilhança.

Os métodos de contagem seguem três passos: Contagem dos sítios sinônimos e não sinônimos nas sequências, seguido da contagem das diferenças (mutações) sinônimas e não sinônimas e cálculo da proporção das diferenças com correção das distâncias. O primeiro método de contagem heurística assumem um modelo simples de mutação de frequências iguais entre os nucleotídeos (Nei e Gojobori, 1986). Um modelo mais realístico foi desenvolvido para acomodar a diferença entre transições e transversões, e também particionando o códon em 3 classes de degeneração diferentes (Li *et al.*, 1996). Entretanto, este modelo assumia a mesma frequências para todos os códons, uma premissa pouco realista.

Outro modelo que acomoda melhor a razão de transição e transversão e assume frequências desiguais de códons, foi desenvolvido por Yang e Nielsen (2000) e é, portanto, mais realista que os modelos anteriores.

Entretanto, o uso da razão dN/dS média para verificar evidência de seleção positiva em uma região mostrou-se ineficaz, pois a seleção natural atua somente em alguns sítios específicos e a maioria dos sítios em uma proteína estão sobre seleção neutra ou negativa. Ao gerar a média dos sítios a razão será, na maioria das vezes, menor ou igual a 1 (Yang, 2007).

Os modelos de máxima verossimilhança assumem que a razão  $\omega$ , para qualquer códon em um gene ou região codificadora, é uma variável aleatória de uma distribuição estatística, permitindo que a função de  $\omega$  varie entre os diferentes códons. Além de acomodar modelos de substituições mais realísticos, os modelos de máxima verossimilhança permitem variação entre os valores de  $\omega$  entre os códons. Esses modelos adaptam um modelo de Markov de substituições de códons ao conjunto de sequências para estimar  $t$  (tempo de divergência),  $\kappa$  (razão de transição e transversão),  $\omega$  e o somatório de  $\pi_j$  (frequência de equilíbrio do códon da posição  $j$ ), de acordo com o modelo proposto por Goldman e Yang (1994). As frequências de equilíbrio dos códons são estimadas a partir do conjunto de dados, enquanto os parâmetros  $t$ ,  $\kappa$  e  $\omega$  são estimados por maximização numérica do logaritmo de verossimilhança  $\ell$ . Com todos os parâmetros estimados, é realizado então o cálculo da proporção de mutações sinônimas sobre sítios sinônimos (dS) e da proporção de mutações não sinônimas sobre sítios não sinônimos (dN), gerando então os valores de dN/dS para cada um dos códons independentemente. Dessa forma, pode-se comparar modelos que permitam variação de  $\omega$  entre 0 e 1, chamados de modelo neutros, com modelos que permitem variação  $0 < \omega < 1$ , chamados de modelos de seleção positiva, realizando-se testes de razão de verossimilhança .

### **1.8 – Ferramentas de Análise.**

As moléculas do complexo principal de histocompatibilidade (*MHC*) são fundamentais para a imunidade celular mediada por células T. O MHC é dividido em duas classes: Classes I e II. As moléculas da classe I são expressas na maioria das células nucleadas onde se ligam aos peptídeos processados pelo sistema de síntese de antígenos. Esses complexos são reconhecidos por linfócitos CD8<sup>+</sup> que, quando ativos, reconhecem e provocam morte nas células do hospedeiro que expressam o MHC ligado à epítomos não próprios, como por exemplo, células infectadas (Maenaka e Jones, 1999). As moléculas da classe II são expressas principalmente por células apresentadoras de antígeno (neutrófilos,

monófilos e células dendríticas) e se ligam a epítomos processados de moléculas extracelulares, processadas via fagocitose, e são apresentados aos linfócitos CD4<sup>+</sup>.

Entretanto, a grande variabilidade apresentada pelos vírus permite uma rápida evolução viral levando o aparecimento de mutações nas regiões reconhecidas pelo sistema imune celular (epítomos). Essas mutações nos epítomos resultam no não pareamento do epítomo com o MHC, o que não induz à resposta imunológica pelo não reconhecimento do epítomo, permitindo o vírus escapar da influência das respostas imunológicas. Esse processo de persistência viral devido às mutações de escape é frequente nos vírus de rápida evolução como HIV e HCV (Goulder e Waticins, 2008).

A avaliação de epítomos em proteínas tem sido um grande desafio para a Biologia Computacional e inúmeras abordagens de soluções já foram propostas. Os métodos de predição são feitos baseados nas informações contidas nos bancos de dados, dos quais se busca reconhecer os padrões de determinados epítomos e ligantes específicos de uma molécula de MHC. O sucesso de um preditor está em um grande conjunto de dados de treinamento representativo. Uma grande variedade de técnicas de aprendizado de máquina são comumente utilizadas no campo da bioinformática imunológica. O método de busca de motivo se baseiam na características do sulco da molécula de MHC, que possui os “bolsos”, os quais aminoácidos específicos se ligam para estabilizar o complexo MHC-epítomo. Esses “bolsos” possuem, dependendo do tipo de MHC, afinidade por resíduos com características particulares, como tamanho da cadeia lateral, hidrofobicidade, e cargas elétricas (Larsen *et al.*, 2007).

Como os sulcos localizam-se, geralmente, nas posições específicas 2, 4, 6 e 9, a sequência dos peptídeos que ligam-se ao MHC possuem aminoácidos específicos nessas posições. Os sítios 2 e 9 são os sítios de ancoramento no MHC e, em alguns tipos, os 4 e 6 são os sítios auxiliares para a ligação. Na maioria dos tipos de MHC da classe I essas posições são específicas e só apresentam um aminoácido correspondente a essa posição (Rammensee *et al.*, 1995; Jardetzky *et al.*, 1996). Já os epítomos de MHC da classe II tem uma tendência de serem mais variáveis nessas posições (chamadas de posições de ancoramento), suportando mais de um aminoácido de características e tamanhos semelhantes, tornando sua predição mais difícil (Wang *et al.*, 2008). Os métodos de matrizes de escore por posição específica também levam em consideração os aminoácidos de ancoramento nas posições dos “bolsos”, entretanto também levam em consideração as outras posições que não são nas de ancoramento. Nessa metodologia os epítomos são convertidos em uma matriz de escore 20 por 9, e os escores representam o perfil dos epítomos. Por esta razão, esse método consegue

considerar as outras posições do epítopo. O escore de cada posição é calculada como a soma de cada contribuição de cada aminoácido para posição, nos peptídeos do banco de dados. Com isso a característica dos aminoácidos em cada posição pode ser derivada. Para obter essa matriz é necessário um grande numero de sequências dos peptídeos que ligam-se, sendo que essa estratégia é necessária para que haja a contribuição máxima de cada aminoácido nas posições (Nielsen *et al.*, 2007). A afinidade do epítopo ao MHC pode ser calculada simplesmente aplicando a matriz obtida à sequência do epítopo, assim podendo determinar a força de ligação. Geralmente, as matrizes são de log-odds, onde os escores são estimados a partir do logaritmo da razão entre a frequência de um determinado aminoácido (em uma posição) e da frequência de fundo desse mesmo aminoácido é obtida a partir de matrizes já definidas, como a matriz Blossum (Henikoff e Henikoff, 1992) e PAM (Dayhoff *et al.*, 1978).

Métodos probabilísticos também são utilizados para a predição de epítopos. Esses métodos são capazes de reconhecer padrões não lineares como, por exemplo, a independência entre as posições dos aminoácidos que se ligam ao sulco do MHC. Outra vantagem desses métodos é a possibilidade de implementação de informação *a priori* dos epítopos. Alguns métodos assumem, além da informação dos epítopos, as ligações entre a proteína TAP (proteína que auxilia no transporte do peptídeo do citoplasma para o lúmen da vesícula onde a molécula do MHC está, no caso da classe I) e também alguns padrões de clivagem específicos do proteassomo, que aumentam o poder preditivo. Além disso, os métodos probabilísticos são treinados com dados de peptídeos ligantes e com peptídeos que não se ligam ao MHC, possibilitando uma melhor caracterização das proteínas que estão sendo testadas e diminuindo assim os falsos positivos (Peters *et al.*, 2005; Larsen *et al.*, 2007).

As Redes Neurais Artificiais são um exemplo de metodologias probabilísticas. Nesse método, ocorre o treinamento feito a partir das informações dos bancos, onde é distribuído em uma rede computacional com uma camada de entrada, onde a informação é introduzida e processada para decidir se mandam o sinal adiante ou cessa nesta camada (perceptron). as camadas escondidas, onde ocorre o processamento da informação em outros níveis onde os sinais são então enviados para a camada de saída, onde a predição é gerada a partir de todos os dados oferecidos ao método (Hardin *et al.*, 2002).

As Cadeias de Markov Escondidas tem sido utilizadas para a predição de epítopos lineares específicos a anticorpos, pois há uma necessidade da caracterização da estrutura desse epítopo. A metodologia é baseada em transições de estados. Após definidos os estados do algoritmo, ele é treinado em um conjunto de dados para calcular as probabilidades de transição entre esses estados. Com isso, os preditores são determinados (Eddy, 1996). O

método caracteriza as posições baseado nas probabilidades calculadas no conjunto de dados de treinamento do algoritmo. Para a predição de epítomos de MHC, as Cadeias de Markov Escondidas são utilizadas devido a facilidade de implementação da informação *a priori* da estrutura do sulco do MHC e de outras características já descritas previamente (Larsen *et al.*, 2007).

Outras metodologias também tem sido utilizadas, como máquinas suportadas a vetor, e alguns métodos de busca de motivo. Entretanto, as metodologias mais robustas para a predição como os métodos de Matriz de Escore por posição Específicas (PSSM), as Cadeias de Markov Escondidas (HMM) e as Redes Neurais Artificiais (ANN), são os mais amplamente utilizados (Peters *et al.*, 2005; Larsen *et al.*, 2007) entretanto, nenhuma das metodologias de predição de epítomo assume a existência de eventos evolutivos nos seus algoritmos e são incapazes de avaliar a seleção natural nas sequências que são avaliadas.

A identificação de epítomos que não possam sofrer mutações sem um grande custo para a replicação viral é a chave para o desenvolvimento de vacinas eficazes contra o vírus (Erickson *et al.*, 2001; Shoukry *et al.*, 2003). As utilizações de programas computacionais para identificação de regiões sob pressão seletiva auxiliariam na identificação de epítomos evolutivamente estáveis. Modelos de avaliação da pressão seletora em regiões codificantes são uma forma de determinar se, em um conjunto de sequências, existem regiões ou posições sobre pressão positiva ou negativa (Yang e Nielsen, 2002). Porém, esses modelos não realizam buscas por assinaturas específicas, sendo necessária a busca não automatizada dos padrões. Além disso, são escassas as informações sobre ferramentas de mapeamento de epítomos já descritos do HCV. As únicas ferramentas disponíveis são os preditores, os quais não são específicos para o vírus e as buscas são feitas em uma sequência de aminoácidos por vez (Lundegaard *et al.*, 2007), impossibilitando o estudo de pressão seletiva, visto que a informação evolutiva do organismo avaliado não está acessível ao algoritmo. Dessa forma, o desenvolvimento de uma ferramenta que permita o mapeamento dos epítomos, juntamente com uma análise de pressão seletiva é necessário para viabilizar essa tarefa.

# Justificativa

## 2 – JUSTIFICATIVA

Estudos recentes demonstraram que o controle da viremia em pacientes infectados pelo HCV está associado com uma forte resposta de células T CD4<sup>+</sup> e CD8<sup>+</sup>, que são específicas para epítomos em indivíduos com carga viral transitória ou indetectável. Nos indivíduos cuja carga viral é alta, o número das células T epítomo específico é baixo (Cooper *et al.*, 1999; Lechner *et al.*, 2000).

Estudos com ratos transgênicos para o HLA A 2.1 demonstraram que ratos imunizados com vírus Canarypox recombinantes com as proteínas não estruturais do HCV NS3-NS4-NS5 apresentaram uma forte resposta imunológica citotóxica e estimulação da secreção de IFN- $\gamma$  (Pancholi *et al.*, 2003).

Nos estudos sobre a evolução da infecção pelo HCV foram observadas sólidas associações entre a persistência da infecção com o surgimento de mutações de escape (Cox *et al.*, 2005). Entretanto, os fatores que promovem o escape viral e a persistência do HCV ainda não estão bem caracterizados.

Como o tratamento contra a infecção pelo HCV com a combinação de Interferon- $\alpha$  e Ribavirina tem taxas de resposta baixas, entre 35 a 40% (McHutchison & Poynard, 1999), e não há uma vacina eficaz na prevenção da infecção, a compreensão de como o vírus se adapta ao tratamento antiviral é necessária para o desenvolvimento de novos tratamentos e na modificação de protocolos atuais. A compreensão da evolução do vírus em pacientes respondedores e não respondedores ao tratamento, juntamente com o perfil de epítomos virais possibilitarão o entendimento dos mecanismos de escape viral frente ao sistema imunológico do hospedeiro, que responde e não responde à terapia, podendo auxiliar no desenvolvimento de tratamentos mais eficazes, assim como no desenvolvimento de uma vacina efetiva para o controle da epidemia.

## Objetivos

### 3 – OBJETIVOS

- Verificar se existem diferenças nos padrões de evolução da proteína NS5A entre os vírus de pacientes respondedores e não respondedores ao tratamento com Interferon- $\alpha$  associado com Ribavirina utilizando o cálculo da razão dN/dS e modelos evolutivos de máxima verossimilhança.
- Mapear os epítomos na proteína NS5A dos vírus nos indivíduos não respondedores e respondedores ao tratamento para verificar se a diferença entre os padrões de evolução da proteína está associada com o aumento da resposta imune contra o HCV.
- Desenvolvimento de uma ferramenta automática para mapeamento de epítomos do HCV e de análise da pressão seletiva e validá-lo com os resultados obtidos das análises prévias.

# material e métodos

## 4 – MATERIAL E MÉTODOS

### 4.1 – Amostragem

Para esse estudo foram utilizados 3 conjuntos de pacientes, divididos devido às suas características distintas. Ao todo foram amostrados 28 pacientes, todos apresentando infecção pelo HCV na forma crônica. 24 pacientes foram tratados com PEG-INF- $\alpha$ -2b e acompanhados durante o tratamento e os outros 6 pacientes sofreram transplante de fígado, pois já apresentavam dano hepático. Dos pacientes tratados, 12 eram infectados com vírus do genótipo 1 e 12 eram infectados com vírus do genótipo 3. Dessa forma, foram formados os grupos dos pacientes infectados com vírus da hepatite C de genótipo 1, dos pacientes infectados com vírus da hepatite C de genótipo 3 e dos pacientes infectados com vírus da hepatite C transplantados. Para cada grupo foram realizadas as análises descritas posteriormente.

#### 4.1.1 – Pacientes infectados por HCV de genótipo 1

As sequências que foram utilizadas no estudo do genótipo 1 foram obtidas de uma amostra de pacientes do Departamento de Hepatologia da UNESP de São José do Rio Preto. Os pacientes portadores de outras doenças hepáticas (infectados com HBV ou outras doenças virais hepatotrópicas, autoimunes, hereditárias ou por abuso de álcool) não foram incluídos no estudo, o que resultou em uma amostra de 12 pacientes infectados pelo HCV dos genótipos 1a (7 pacientes) e 1b (5 pacientes). As amostras foram obtidas após aprovação do comitê de ética da UNESP.

As amostras foram coletadas antes do início do tratamento com PEG-INF- $\alpha$ -2b subcutâneo e de 600 até 1200 mg ribavirina por via oral, de acordo com o protocolo médico de tratamento. Os pacientes foram acompanhados durante um ano, que é o período de tratamento, e 6 meses após o término do tratamento.

Em uma primeira abordagem, os pacientes que apresentaram ausência de RNA viral 6 meses após o término do tratamento foram classificados como Respondedores ao Tratamento (RT). Consequentemente, os que apresentaram presença contínua do RNA viral durante o tratamento foram classificados como Não respondedores (NR) ou que demonstraram ausência de RNA viral durante o tratamento, porém apresentaram RNA viral após o fim do tratamento foram classificados como Não Respondedores de Final de Tratamento (NRFT) como descrito por Jardim *et al.* (2009) e Bittar *et al.* (2010). Para cada paciente foram realizadas 15

clonagens, com exceção de 2 pacientes dos quais foram obtidos 14 clones. Toda região NS5A foi amplificada e sequenciada (1344 nucleotídeos), entretanto, não foi possível amplificar a região NS5A dos clones virais de um paciente do sexo masculino infectado por genótipo 1a, resultando em uma amostra final de 11 pacientes sequenciados, dos quais foram analisados 163 clones no estudo do pré tratamento no total.

Em uma segunda abordagem, os pacientes que apresentaram presença contínua do RNA viral durante o tratamento e os que demonstraram ausência de RNA viral durante o tratamento, porém apresentaram RNA viral após o fim do tratamento foram classificados em um único grupo de Não Respondedores (NR) como descrito por de Queiróz *et al.* (2010). Os grupos RT e NR demonstraram média de idade similar, 39,6 e 39,7 anos respectivamente, sendo a maioria do sexo masculino (10 pacientes do sexo masculino e 2 do sexo feminino).

#### **4.1.2 – Pacientes infectados por HCV de genótipo 3**

As sequências utilizadas para o estudo do genótipo também são de pacientes do Departamento de Hepatologia da UNESP de São José do Rio Preto. Os mesmos critérios de exclusão aplicados aos pacientes infectados pelo genótipo 1 foram utilizados nessa amostra, totalizando 12 pacientes infectados com HCV do genótipo 3.

Novamente, os pacientes foram acompanhados durante um ano e os que apresentaram ausência de RNA viral 6 meses após o término do tratamento foram classificados como Respondedores ao Tratamento (RT), os que apresentaram presença contínua do RNA viral durante o tratamento foram classificados como Não respondedores (NR) ou que demonstraram ausência de RNA viral durante o tratamento porém apresentaram RNA viral após o fim do tratamento foram classificados como Não Respondedores de Final de Tratamento (NRFT) como descrito por Jardim *et al.* (2009) e Bittar *et al.* (2009).

Para cada paciente toda a região NS5A foi amplificada e sequenciada em 15 clones das amostras coletadas antes do início do tratamento com PEG-INF- $\alpha$ -2b subcutâneo e ribavirina por via oral, com exceção de um paciente de quem foram gerados 14 clones. No total, foram geradas 179 sequências a partir de clones dos pacientes estudados.

#### **4.1.3 – Pacientes transplantados Região NS3**

As sequências utilizadas no estudo das diferenças de epítomos entre os genótipos 1 e 3 foram obtidas de 6 pacientes, submetidos a transplante de fígado devido à cirrose hepática, do

Departamento de Hepatologia da Universidade Louis Pasteur de Strasburgo. Destes, 3 eram infectados por HCV genótipo 1 e o restante pelo genótipo 3. A carga viral dos pacientes variou de 0,7 até  $9 \times 10^5$  IU/ml, antes do transplante de fígado, e de 5,9 até  $7,5 \times 10^5$  IU/ml, três meses após o transplante. Os pacientes não receberam tratamento antiviral até o 1º mês após o transplante.

## **4.2 – Análise das sequências dos Pacientes**

### **4.2.1 – Pacientes infectados por HCV de genótipo 1**

Os alinhamentos, a partir do quadro aberto de leitura das sequências ortólogas obtidas, foram realizados utilizando-se o programa Muscle (Edgar, 2004) com as penalidades de abertura de lacuna igual a 10 e extensão de lacuna igual a 0,1. A edição visual dos alinhamentos, para verificar o alinhamento automático e editar o conjunto de dados, foram realizadas utilizando os programas GeneDoc 2.6 (Nicholas *et al.*, 1997) e BioEdit 5.0.9 (Hall, 1999), essa metodologia foi aplicada para as abordagens de três Grupos (Respondedor, Não respondedor e Respondedor ao Final do Tratamento) e de dois grupos (Respondedor e Não Respondedor).

O numero de mutações de nucleotídeos e de aminoácidos, juntamente com a distância genética (distância-p) foram estimados utilizando-se o programa MEGA 4.0 (Tamura *et al.*, 2007). O programa SNAP – Synonymous Non-synonymous Analysis Program – foi utilizado para calcular os valores da proporção de mutações não sinônimas por sítios não sinônimos (dN) e os valores da proporção de mutações sinônimas por sítios sinônimos (dS), por comparação par a par de todas as sequências da região NS5A. Os valores de dS e dN foram utilizados para gerar o valor médio de dN/dS da região (Korber, 2000). Além desses parâmetros, foi estimada a diversidade genética (Pi), utilizando-se o programa DNAsp 4.5 (Rozas *et al.*, 2003). Essas metodologias foram aplicadas para a abordagem de três grupos (Respondedor, Não respondedor total e Não respondedor de final de tratamento).

Após esta etapa foi realizada a tradução conceitual dos alinhamentos gerados utilizando-se o programa GeneDoc 2.6 (Nicholas *et al.*, 1997) no primeiro quadro aberto de leitura, em que as frequências dos epítomos de células CD8<sup>+</sup> contidos no HCV Immunology database (Yusim *et al.*, 2005) foram calculadas. Também foi realizada a predição de sítios potenciais de modificações pós traducionais utilizando a ferramenta Prosite (Falquet *et al.*, 2002). Essa metodologia foi aplicada para a abordagem de dois Grupos (Respondedor e Não

respondedor).

A razão das mutações sinônimas e não sinônimas (razão dN/dS) entre as sequências codificantes de proteínas foi estimada utilizando-se o programa PAML versão 4 (Yang 2007), com duas abordagens: Utilizando o modelo evolutivo de máxima verossimilhança M0 (uma razão) para a região NS5A inteira, que estima o valor da razão que maximiza a função de verossimilhança, e o modelo par a par YN00 somente para as regiões codificadoras correspondentes aos epítomos (Yang e Nielsen, 2000), que estima a razão dN/dS calculando os valores da proporção de mutações não sinônimas por sítios não sinônimos (dN) e os valores da proporção de mutações sinônimas por sítios sinônimos (dS) diretamente da comparação entre as sequências do conjunto de dados. Os modelos de verossimilhança foram utilizados para determinar seleção positiva sem utilizar o valor médio da razão dN/dS através da árvore filogenética. Os modelos de máxima verossimilhança baseados em substituições de códons M1a (neutro), M2a (seleção), M7(beta) e M8 (beta e ômega) foram utilizados para testar evidência de pressão seletiva positiva nos conjuntos de dados por comparação dos valores de verossimilhança (indiretamente pela comparação dos valores de Ln da verossimilhança) dos modelos M1a contra M2a e M7 contra M8 utilizando o teste de razão de verossimilhança (Yang 2007). O método "Bayes Empirical Bayes" foi aplicado quando o teste de razão de verossimilhança sugeriu evidência de pressão positiva para indicar qual códon está sobre pressão de seleção positiva (Yang, Wong e Nielsen, 2005). Essas metodologias foram aplicadas para a abordagem de dois grupos.

A diferença estatística dos valores de dN, dS, dN/dS, Mutações de nucleotídeos e de aminoácidos da região NS5A inteira, medida pelo programa SNAP, foi determinada utilizando análise de variância (ANOVA) com teste de Tukey para comparação entre os grupos. A comparação entre os grupos das variáveis distância genética e diversidade genética foram realizadas com o teste não paramétrico de Kruskal-Wallis.

As diferenças estatísticas nos parâmetros dN, dS e dN/dS somente da região codificadora de epítomos dos diferentes grupos foi determinada utilizando-se o teste não paramétrico para valores numéricos de Mann-Whitney. A comparação de presença/ausência dos epítomos entre os grupos foi realizada utilizando-se o teste G. Nós aplicamos a correção conservadora de Bonferroni para múltiplas comparações na análise de presença/ausência dos epítomos.

O teste de razão de verossimilhança, que consiste no cálculo do dobro da diferença dos valores de logaritmo natural de verossimilhança dos modelos testados e na comparação desse valor com a distribuição de  $\chi^2$  para definição do valor de  $p$  entre as hipóteses testadas,

foi realizado para comparação dos modelos de máxima verossimilhança relacionados (M1 neutro vs M2 seleção positiva); M7 beta vs M8 beta e omega), com valores críticos de 5,99 e 9,21 para os níveis de significância 5% e 1% , respectivamente.

#### **4.2.2 – Pacientes infectados por HCV de genótipo 3**

Os alinhamentos múltiplo das sequências obtidas de cada grupo de pacientes (Respondedor, Não respondedor e Respondedor ao Final do Tratamento), foram realizados utilizando-se o programa MUSCLE (Edgar, 2004) com as penalidades descritas acima. A edição dos alinhamentos, para verificar o alinhamento automático e editar o conjunto de dados, foi realizada utilizando-se os programas GeneDoc 2.6 (Nicholas *et al.*, 1997) e BioEdit 5.0.9 (Hall, 1999).

Após gerar o alinhamento múltiplo para cada grupo, o número de mutações de nucleotídeos e de aminoácidos, e a distância genética (distância-p) foram estimados utilizando o programa MEGA 4.0 (Tamura *et al.*, 2007). O programa SNAP – Synonymous Non-synonymous Analysis Program – foi utilizado para calcular os valores da proporção de mutações não sinônimas por sítios não sinônimos (dN) e os valores da proporção de mutações sinônimas por sítios sinônimos (dS) por comparação par a par de cada sequência para a região NS5A inteira para gerar o valor da razão dN/dS (Korber, 2000).

A diferença estatística dos valores de dN, dS, dN/dS, Mutações de nucleotídeos e de aminoácidos da região NS5A inteira, mensurado pelo programa SNAP, entre os grupos foi determinada utilizando-se análise de variância (ANOVA) com teste de Tukey para comparação múltipla entre os grupos.

#### **4.2.3 – Pacientes transplantados Região NS3**

Os alinhamentos múltiplos das sequências foram realizados utilizando-se o programa ClustalX 1.8 (Thompson *et al.*, 1997). Após o alinhamento, foi utilizado o programa MEGA 4.0 (Tamura *et al.*, 2007) para estimar a distância genética (p). Para calcular os valores de dN, dS e o valor da razão dN/dS foi utilizado o programa SNAP – Synonymous Non-synonymous Analysis Program – (Korber, 2000). A frequência dos epítomos contidos no *HCV Immunology database* (Yusim *et al.*, 2005) nas sequências foi avaliada utilizando uma versão prévia da ferramenta de análise REEvolution e os resultados foram confirmados por avaliação visual com o programa GeneDoc 2.6 (Nicholas *et al.*, 1997).

### 4.3 – Ferramenta de Análise REEvolution

A ferramenta de mapeamento dos epítomos do HCV foi desenvolvida na linguagem de programação *python*. As coordenadas de regiões do genoma viral são de acordo com a sequência de referência H77 (número de acesso M67463) e os epítomos contidos no *HCV Immunology database* (Yusim *et al.*, 2005) para o vírus da hepatite C. Para o vírus HIV foram utilizados as coordenadas de regiões em relação a sequência de referência HXB2 (numero de acesso K03455) e os epítomos contidos na *HIV Immunology database* (Korber, 2007). A ferramenta também avalia sequências do vírus da imunodeficiência humana, pois esse vírus também apresenta elevada taxa de mutação além da interação vírus-hospedeiro e resposta imunológica serem semelhante ao do vírus da hepatite C (Brumme *et al.*, 2009). Além disso, o banco de dados de epítomos específicos para o HIV utilizado é padronizado da mesma forma que o banco de dados de epítomos específicos para o vírus da hepatite C, o que permitiu sua implementação na ferramenta.

Devido a observação na análise dos pacientes transplantados, nós também utilizamos para a análise do HCV os epítomos contidos no IEDB (*Immune Epitope Database*, Vita *et al.*, 2010), para aumentar o número de epítomos analisados. O esquema de análise foi dividido em 3 etapas:

#### 1 – Mapeamento dos epítomos:

Após receber e validar os dados informados pelo usuário, a ferramenta realiza a tradução do alinhamento múltiplo das sequências codificantes, no primeiro quadro aberto de leitura, e gera o consenso das sequências de aminoácidos para que a posição em relação à referência H77 ou em relação à referência HXB2 seja estimada utilizando-se o programa "glsearch" do pacote FASTA versão 35.4.7 (Pearson e Lippman, 1988). Após essa etapa, o programa seleciona os epítomos do banco de dados relacionado a posição das sequências teste do vírus analisado e realiza o mapeamento dos epítomos que estão 100% conservados e os que têm até 2 resíduos diferentes, localizando-os nas sequências do usuário, gerando o posicionamento dos epítomos em relação à sequência de referência, assim como nas sequências testadas, permitindo ao usuário comparar com os resultados da análise de pressão seletiva e disponibilizando a localização exata dos epítomos.

#### 2 – Teste para verificar a evidência de seleção natural:

Após esta etapa a ferramenta realiza a análise com os modelos evolutivos de máxima verossimilhança, em que uma árvore será gerada utilizando o critério da máxima verossimilhança e os valores de logaritmo natural da verossimilhança de cada modelo são

calculados. O teste de razão de verossimilhança é realizado, comparando os modelos propostos por Yang (2007), para verificar quais dos modelos é os mais adequados para o conjunto de dados, indicando se há evidência de pressão seletiva ou neutralidade. O modelo M0 é utilizado para calcular o valor da razão dN/dS da região inteira do alinhamento. Dois modelos de evolução neutra (M1 e M7) e dois modelos de seleção positiva (M2 e M8) são utilizados para verificar evidência de seleção positiva por comparação do resultado do teste de razão de verossimilhança com a distribuição de  $\chi^2$ . O resultado da estatística indicará evidência, se o valor de  $p$  da comparação dos modelos for significativo.

### 3 – Análise com *Bayes Empirical Bayes*:

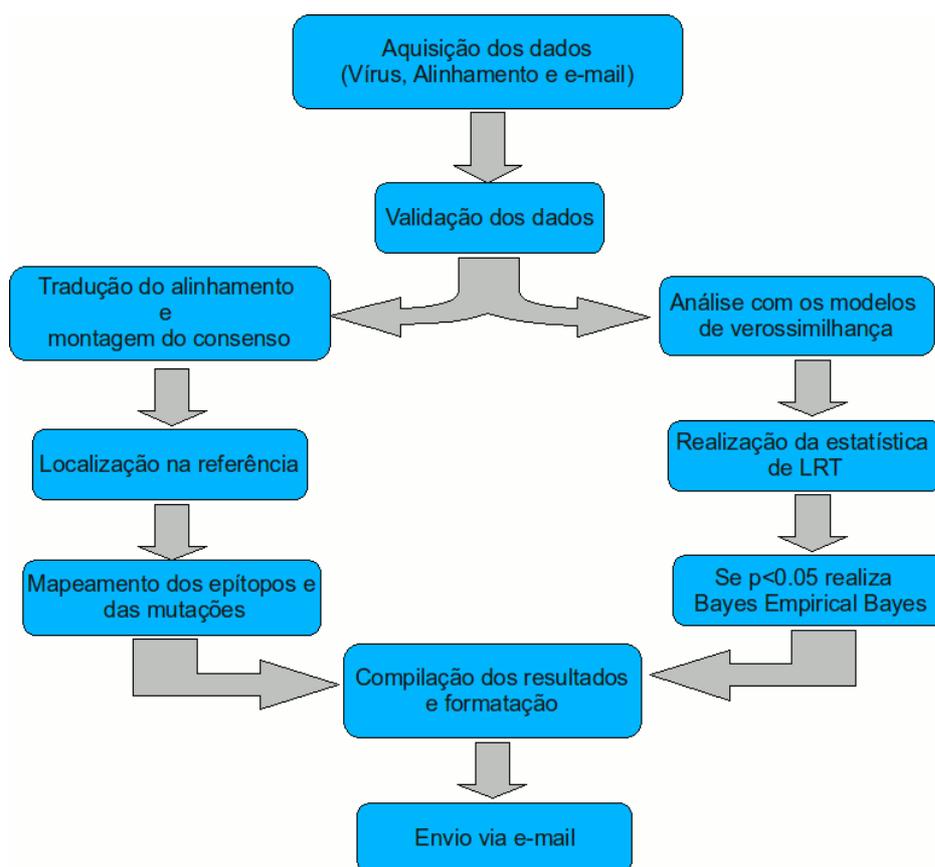
Com evidência de pressão seletiva, a ferramenta gera o resultado da análise *Bayes Empirical Bayes*, que indicará a probabilidade posterior de cada códon pertencer a classe de sítios com pressão positiva de acordo com os modelos 2A e 8 e o valor da razão dN/dS dos respectivos códons. Essa análise permite identificar os códons que estão sobre influência de seleção positiva, onde são gerados os valores da razão dN/dS e a probabilidade posterior do códon pertencer a classe de seleção positiva, caso a razão seja maior que 1 e a probabilidade posterior ( $pr > 0,95$ ).

O resultado será compilado e enviado em formato HTML ao e-mail informado pelo usuário com a frequência dos epítomos no alinhamento, as informações contidas no banco de dados, as frequências do mapeamento, a posição do epítomo na sequência teste, o valor da razão dN/dS para a região, o resultado do teste de razão de verossimilhança e, quando o teste indicar pressão seletiva, os aminoácidos selecionados positivamente. A ferramenta está disponibilizada como um *webtool open source* hospedado em um servidor próprio no Instituto Butantan e no Instituto de Biociências da Universidade de São Paulo, no endereço <http://bioinfo.ib.usp.br/revolver/index.html>.

Para avaliar o desempenho foram realizadas análises com a ferramenta, em que foram utilizados 7 conjuntos de dados com número de sequências diferentes (com 10 sequências, 14, 18, 25, 33, 45 e 60, com 1344 nucleotídeos, que formam uma progressão geométrica) para verificar o tempo necessário para a ferramenta completar as análises.

Com os valores de tempo necessário para cada um dos conjuntos foi realizada a comparação dos valores de F da regressão linear, exponencial e polinomial. Além disso, foram realizadas análises com 3 tamanhos de sequência diferentes (99 nucleotídeos, 498 e 999) com número de sequências diferentes (10, 20, 30, 40 e 50) para verificar o desempenho com relação ao tamanho e número das sequências.

### 4.3.1 – Fluxograma da ferramenta REEvolution.



Em 2007, o NIH parou de fomentar o *HCV Immunology database* e seu banco de dados parou de ser atualizado no final de 2007, com ao todo 383 epítomos específicos para o vírus da hepatite C. Em comparação ao *HIV Immunology database*, que possui 1290 epítomos, o banco de dados de epítomos para o HCV relativamente pequeno. Além disso, o banco possui um viés para o genótipo 3 (Carvalho-Mello *et al.*, 2009), sendo necessárias diversas atualizações no banco.

Para o desenvolvimento da ferramenta nós utilizamos o banco de dados não curado IEDB como fonte dos epítomos para o mapeamento. Estavam contidos no banco 1735 epítomos relacionados com resposta associada aos linfócitos T citotóxicos e ao MHC da classe I, entretanto, haviam epítomos relacionados com resposta negativa (epítomos ligantes ao MHC, porém sem gerar resposta imunológica) avaliados com ELISPOT e liberação de interferon quando apresentados. Após essa avaliação, foram excluídos 435 epítomos que não apresentavam liberação de interferon por ELISPOT. Dos 1300 epítomos restantes, foram excluídos os epítomos que não eram associados aos HLAs humanos da classe I A, B e Cw, excluindo os HLAs não humanos e da classe II. Além disso, os epítomos com tamanho maior

que 13 resíduos, associados com classe II, e aqueles redundantes foram eliminados do conjunto de dados. Após a avaliação foram excluídos 1210 epítomos, totalizando 525 epítomos, restritos à 48 alelos de MHC da classe I, não redundantes no banco de dados final, aumentando em 145 epítomos em relação ao *HCV Immunology database*. O banco de dados final encontra-se disponível para download no site da ferramenta.

# Resultados

## 5 – RESULTADOS

### 5.1 – Pacientes infectados por HCV de genótipo 1

Os resultados das análises utilizando a abordagem de três grupos de pacientes infectados por vírus do genótipo 1 foram publicados em julho de 2009 (Jardim *et al.*, 2009). As características demográficas dos grupos de pacientes foram: Idade 42,5 anos para o grupo Não respondedor (NR), 38,75 para o grupo respondedor (RT) e de 36,75 para o grupo Não Respondedor de Final de Tratamento (NRFT). A maioria dos pacientes era predominantemente do sexo masculino (83,3%). Nos resultados não foram observadas diferenças significantes nas características demográficas (idade, peso, altura, região geográfica) entre os grupos de pacientes. As análises de distância genética apresentaram diferenças estatisticamente significantes com valores de 0,134 para o grupo RT, 0,123 para o grupo NR e de 0,117 para o grupo NRFT. Os resultados da razão dN/dS da região NS5A completa mostrou que a razão era maior no grupo RFT, com valor de 0,067, em comparação aos grupos RT e NR, valores 0,018 e 0,013 respectivamente, estatisticamente significante. Porém, todos os grupos apresentam valores que sugerem seleção purificadora, visto que a razão dN/dS é menor que 1. Apesar dos valores estatisticamente significantes encontrados, nenhum padrão foi observado na região NS5A que pudesse ser associado com resposta ao tratamento.

Após essa abordagem de três grupos, nós utilizamos apenas estratificação em dois grupos, em que os pacientes dos grupos Não respondedor e Não Respondedor de Final de Tratamento foram analisados apenas como Não respondedores. Com os grupos de pacientes redefinidos foi realizado o mapeamento dos epítomos de células CD8+ específicos da região NS5A dos vírus genótipo 1 em todas as sequências de aminoácidos traduzidos do grupo de pacientes respondedores (n=44 sequências) e do grupo não respondedor (n=119). No total, 21 epítomos da região NS5A foram utilizados (de Queiróz *et al.*, 2010).

Três epítomos não foram encontrados nas sequências de ambos os grupos (epítomos: VLDSFKTWL, LPKLPGVPF e QLSAPSLKATCTANH) e um dos epítomos não foi encontrado no grupo respondedor e demonstrou frequência baixa no grupo não respondedor (epítomo KFPPALPIW, com 5,04%). Foram observados cinco epítomos com frequência maior no grupo não respondedor em comparação ao grupo respondedor e oito epítomos com frequência superior no grupo respondedor, entretanto sem suporte estatístico. Dos 21 epítomos testados, quatro apresentaram diferenças estatisticamente significantes entre os dois grupos de

resposta ao tratamento, utilizando o teste G. Assim, três desses epítomos demonstraram frequência superior no grupo respondedor (epítomos: VLSDFKTWL, SVASSASQLSA e ARGSPPSVASSASQLSAPS), sendo que o epítomo VLSDFKTWL mostrou 100% de frequência no grupo respondedor e apenas 48,5% no grupo que não responde ao tratamento. Entretanto, o epítomo LTDPSHITA apresentou frequência maior no grupo não respondedor, com 98% em comparação a 65% do grupo respondedor.

Todos os epítomos que apresentaram diferenças estatisticamente significantes entre os dois grupos são identificados como ligantes ao HLA-A (alelos A1 e A2). Os resultados do mapeamento dos epítomos estão apresentados na tabela 1.

Tabela1: Frequência dos epítomos nas sequências da região NS5A do HCV dos grupos de pacientes (genótipo 1 a/b).

EPITOPO	Alelo	RT	NR
VLSDFKTWL	A2	100	48,73 *
VLDSFKTWL	A2	0	0
SDFKTWLKA	Patr-B*1601	65,9	47,89
DFKTWLQSKL	ND	34,09	50,42
LPKLPGVPF	B35	0	0
MWSGTFPINAY	Patr-A*0601	63,63	49,57
LLREEVSFRV	A2	34,09	49,57
HEYPVGSQI	B60, H-2k	63,63	48,73
PCEPEPDVAVL	B35	31,81	47,05
EPEPDVAVL	B35	31,81	47,05
LTDPSHITA	A1	65,9	98,31 *
ARGSPPSVASSASQLSAPS	A1/A2	65,9	25,21 *
SVASSASQLSA	A1/A2	65,9	25,21 *
QLSAPSLKATCTANH	ND	0	0
NHDSPDAEL	B38	61,36	49,57
SPDAELIEANL	A2	65,9	49,57
ELIEANLLW	A25	65,9	49,57
ILDSFDPLV	A2	31,81	10,92
LDSFDPLVA	Patr-B*0401	65,9	47,89
REISVPAEIL	B60	31,81	49,57
KFPPALPIW	A*2402	0	5,04

ND, Não disponível;

Nota – \*  $p < 0.05$  (corrigido com o procedimento de Bonferroni);

Todas as frequências expressas em porcentagem;

Para avaliar a variação dos aminoácidos e determinar os possíveis códons selecionados positivamente, nós utilizamos modelos de máxima verossimilhança cujos resultados estão apresentados na tabela 2. Os valores da razão dN/dS da região NS5A dos grupos respondedor, não respondedor e os dois grupos juntos, calculadas pelo modelo M0 não mostraram diferença significativa. Na comparação do primeiro par de modelos, M1a(neutro) e M2(seleção positiva), não foi observada significância estatística no teste de razão de

verossimilhança, entretanto, no segundo par de modelos, M7(beta) e M8(beta e ômega), o teste mostrou que o Modelo M8 estava selecionado, indicando evidência de variação na pressão seletiva em códons ao longo da região codificadora da proteína NS5A tanto no grupo que não responde ao tratamento como quando os dois grupos são analisados juntos.

Tabela 2: Resultados da análise de pressão seletiva entre os grupos de pacientes.

Resposta	Razão dN/dS	M2a vs M1a	M8 vs M7	AA selecionados positivamente (posição)
RT	0,09	0,367	4,058	ND
NR	0,09	0,023	13,376*	L(2130), P(2319)
Ambos	0,1	0,001	15,49*	L(2130), P(2319)

ND – não detectado;

AA – aminoácido;

Nota - \*  $p < 0.01$  no teste de razão de verossimilhança;

Com a evidência de seleção positiva, indicada pelo teste de razão de verossimilhança, o método *Bayes Empirical Bayes* (BEB) foi aplicado para calcular a probabilidade *a posteriori* de cada códon pertencer a classe de sítios selecionados positivamente de acordo com o modelo M8, indicando assim que o aminoácido L (lisina), na posição 2130 (de acordo com a sequência referência H77, número de acesso: AF009606), e o aminoácido P (prolina), na posição 2319, estão selecionados positivamente com probabilidade posterior maior que 95%.

Para os quatro epítomos que apresentaram diferenças entre os grupos, com suporte estatístico no teste G, foram calculados o valor de dN/dS, dS e dN, utilizando modelos par a par YN00 (tabela 3). Três epítomos apresentaram razão dN/dS maior no grupo não respondedor e um apresentou razão maior no grupo respondedor, todos com suporte estatístico pelo teste não paramétrico de Mann-Whitney, sendo que os epítomos que apresentaram valores superiores foram aqueles que apresentaram frequências inferiores quando comparados ao outro grupo. No cálculo das mutações não sinônimas dois epítomos demonstraram valores superiores no grupo não respondedor e um apresentou valor maior no grupo respondedor, com suporte estatístico. Na comparação das mutações sinônimas foi observado que 3 epítomos obtiveram valores superiores no grupo não respondedor, com suporte estatístico.

Tabela 3: Valores da razão dN/dS, dN e dS .

Epítopo	Resposta	dN/dS	dN	dS
VLSDFKTWL	RT	0,03	0,02	0,76
	NR	0,04*	0,03*	0,91*
LTDPSHITA	RT	0,018*	0,026*	0,79
	NR	0,001	0,001	0,86*
ARGSPPSVASSASQLSAPS	RT	0,02	0,01	0,29
	NR	0,09*	0,01	0,34*
SVASSASQLSA	RT	0,06	0,02	0,25
	NR	0,12*	0,03*	0,25

Nota – \*  $p < 0.05$  no teste de Mann-Whitney;

De acordo com o resultado do mapeamento dos epítomos, nenhuma variação no epítopo VLSDFKTWL foi observada no grupo respondedor, porém no grupo que não responde ao tratamento foi encontrada modificação na 3<sup>a</sup> posição do epítopo, onde uma serina foi substituída por treonina com 37,8% de frequência e por alanina com 12,6%. A análise das sequências do epítopo LTDPSHITA demonstrou variação na 2<sup>a</sup> posição, onde treonina foi substituída por serina em 34% das sequências do grupo respondedor. As variações no grupo não respondedor foram menores que 2%. O epítopo ARGSPPSVASSASQLSAPS apresentou variação na 8<sup>a</sup> posição, onde valina foi substituída por leucina em 33.9% no grupo respondedor e em 63% no grupo não respondedor. Nessa mesma posição, no grupo não respondedor, valina foi substituída por metionina em 12,6% das sequências. No epítopo SVASSASQLSA foi observada variação na 2<sup>a</sup> posição onde valina foi substituída por leucina em 33.9% no grupo respondedor e em 63% no grupo não respondedor e por metionina no ultimo grupo em 12,6% das sequências. Nenhuma das variações nos epítomos apresentou evidência de pressão seletiva positiva na análise com o método *Bayes Empirical Bayes* BEB. Os resultados com as mutações encontradas nos epítomos estão expressos na tabela 4.

## 5.2 – Pacientes infectados por HCV de genótipo 3

Os resultados da análise das sequências dos pacientes infectados pelo HCV do genótipo 3 (Bittar *et al.*, 2010) mostram uma variação de carga viral entre log 5,76 e log 7,03, esperadas para amostras de pré tratamento. A média da carga viral no grupo RT foi de 5.578.634, no grupo NR foi de 2.348.652 e no grupo NRFT foi de 2.422.973. Entretanto não foi possível associar a carga viral com a resposta ao tratamento.

Tabela 4: Mutações observadas nas regiões dos epítomos com significância estatística entre os grupos de pacientes.

	VLSDFKTWL	LTDP SHITA	ARGSPPSVASSSASQLSAPS	SVASSSASQLSA
RT	.....	.S.....	.....L..... .....L...L..... .G.....L.....	.L..... .L...L.....
NR	..T..... ..A.....	...S..... ...T.....	.....M..... .....L.....	.M..... .L.....

Apesar do grupo RT ter demonstrado valores maiores de distância genética, número de mutações de nucleotídeos e de aminoácidos em relação aos grupos NR e RFT, não foram encontradas diferenças estatisticamente significantes. O grupo RT também apresentou valores de dN e de dS maiores em relação aos outros grupos, com diferença significativa, entretanto o grupo NR foi o que apresentou valores da razão dN/dS maiores, com diferença estatisticamente significativa. Apesar das diferenças encontradas, todos os grupos apresentam evidência de seleção purificadora, visto que os valores da razão dN/dS foram menores que 1.

Assim como no trabalho com o genótipo 1 (Jardim *et al.*, 2009), não foram encontrados padrões que pudessem ser associados com a resposta ao tratamento, apesar das diferenças estatisticamente significantes entre os grupos.

### 5.3 – Pacientes Transplantados Região NS3

Além dos trabalhos já citados, também fizemos análises de mapeamento de epítomos utilizando a proteína NS3 de pacientes transplantados, infectados com vírus de genótipos 1 e 3 (Carvalho-Mello *et al.*, 2009), em que a frequência dos epítomos foi calculada utilizando uma versão prévia da ferramenta de mapeamento e confirmado por análise visual. O resultado demonstrou que os valores de dS eram maiores que os valores de dN, resultando numa razão dN/dS menor que 1, indicando evidência de seleção purificadora nos vírus dos pacientes transplantados. O resultado do mapeamento dos epítomos revelou que, dos 30 epítomos estudados, apenas 3 foram encontrados tanto nos vírus de genótipo 3 quanto nos vírus de genótipo 1 com 100% de frequência (tabela 5). No total, 10 epítomos foram mapeados apenas nas sequências de genótipo 1, com frequências variando de 16 à 100%. Não foram mapeados 17 epítomos relacionados com a região NS3 nas sequências dos dois genótipos.

Tabela 5: Frequência dos epítomos nas sequências da região NS3 dos pacientes transplantados infectados pelos genótipos 1 e 3.

Localização referência H77	Epítopo (Genótipo)	Frequência	
		Genótipo 1	Genótipo 3
1031–1039	AYSQQTRGL(1b)	61,11	0
1038–1047	GLLGCIITSL(1b)	88,89	0
1055–1069	VEGEVQIVSTAAQTF(1a)	0	0
1067–1081	QTFLATCINGVCWTV(1a)	0	0
1071–1090	ATCINGVCWTVYHGAGTRTI(1a)	0	0
1073–1081	CINGVCWTV(1a)	16,67	0
ND	QLVALGINAV(1a)	0	0
ND	CVNGVCWTV(2b)	83,33	0
ND	CVVGVCWTA(nd)	0	0
ND	SVNGVCWTV(nd)	0	0
1085–1100	AGTRTIASPKGPVQM(1a)	0	0
1087–1101	TRTIASPKGPVIQMY(1a)	0	0
1100–1107	MYTNVDQD(nd)	100	100
1007–1121	DLVGWPAPQGSRLT(1a)	0	0
1123–1131	CTCGSSDLY(1a)	100	0
1131–1139	YLVTRHADV(nd)	100	0
1169–1177	LLCPAGHAV(1a)	0	0
ND	LLCPSGHVV(nd)	0	0
1171–1190	CPAGHAVGIFRAAVCTRGVA(1a)	0	0
1175–1183	HAVGLFFRAA(1a)	0	0
1243–1252	ATLGFAYM(1b)	83,33	0
1261–1270	TLGFAYMSK(1a)	83,33	0
1262–1270	LGFGAYMSK(HCV-H)	83,33	0
1265–1274	GAYMSKAHGV(1b)	88,89	0
1284–1298	TITGAPVTYSTYK(1a)	0	0
1287–1296	TGSPITYSTY(1b)	0	0
ND	TGAPVTYSTY(1b)	0	0
1292–1300	TYSTYKFL(1b)	100	100
1300–1310	LADGGCSGGAY(1a)	100	100
1318–1327	CHAQDATTVL(3a)	0	0

ND – não determinado;

#### 5.4 – Ferramenta de Análise

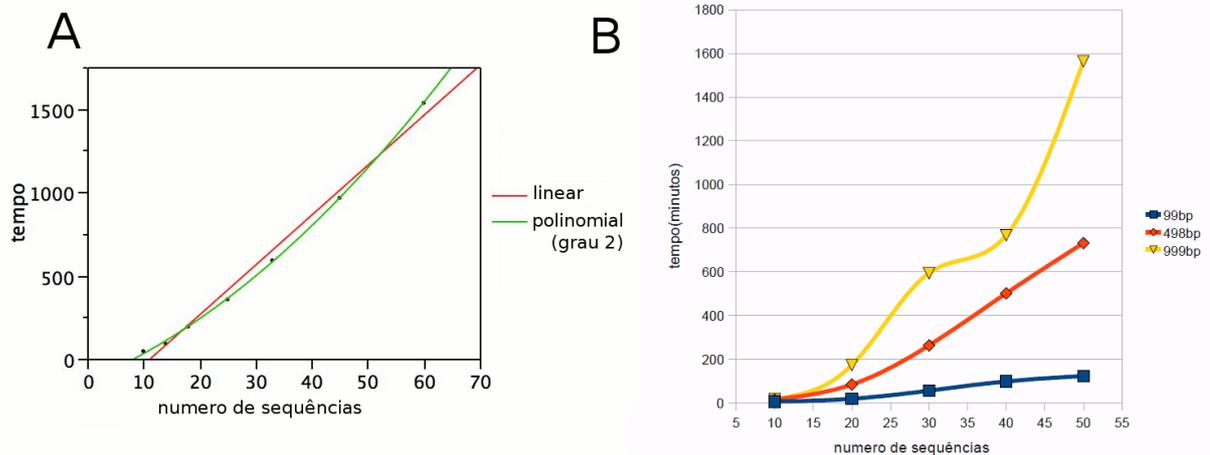
Na avaliação dos epítomos realizada com as sequências dos pacientes transplantados e confirmada com avaliação manual a ferramenta foi capaz de mapear os mesmos epítomos da tabela 5, além de mapear as mutações nas regiões correspondentes aos mesmos. A confirmação manual revelou que tanto o mapeamento dos epítomos quanto os das mutações foram realizados corretamente, com 100% de acerto da ferramenta, porém revelou que o banco de dados do HCV *Immunology database* indicava uma quantidade pequena de epítomos específicos para o genótipo 3. Para minimizar esse efeito, nós também utilizamos os epítomos contidos na IEDB, resultando no aumento de 380 para 530 epítomos específicos para o HCV.

A análise do desempenho da ferramenta, através da comparação dos Fs da regressão linear, exponencial e polinomial (Linear F= 40, exponencial F= 407 e polinomial F=4928),

mostrou que a ordem do algoritmo é quadrática, ou seja, é proporcional ao quadrado de um termo que contém o número e o tamanho das sequências (figura 4 A),  $O(n^2)$ . O tempo necessário para avaliar N sequências com 1344 nucleotídeos é obtido pela equação:

$$\text{Tempo} = -331,3422 + 27,534199 N + 0,2290466 (N - 29,2857)^2 \text{ min}$$

Também foi notado que a ferramenta é mais eficiente com sequências de tamanho pequeno, com os quais o tempo necessário para analisar 50 sequências com 99 nucleotídeos foi de 2 horas e 4 minutos. Para analisar o mesmo número de sequências, porém com 999 nucleotídeos foram necessárias mais de 26 horas (figura 4 B) utilizando uma CPU com processador Intel™ Core Duo Quad Q8400S com 4 núcleos e frequência de 2.66 Ghz, com 3 gigabytes de RAM. O grande tempo necessário para análise é devido principalmente ao uso dos modelos de evolução. O mapeamento dos epítomos é realizado em poucos minutos.



**Figura 4:** Gráficos da análise de desempenho da ferramenta: A- Gráfico linear e polinomial do tempo (em minutos) para determinação da ordem do algoritmo. B- Gráfico dos tempos para os diferentes tamanhos e numero de sequências.

# Discussão

## 6 – DISCUSSÃO

Em nossos trabalhos de análise prévia dos pacientes infectados pelo HCV dos genótipos 1 e 3 não encontramos relação significativa entre as variáveis analisadas que permitisse correlação entre a variabilidade do HCV e o resultado do tratamento combinado de interferon e ribavirina, embora sugerissem que o grupo que responde ao tratamento mostrou valores de mutações nucleotídica e de aminoácidos e de distância genética maiores que o grupo que não respondeu na região NS5A inteira e variando nas sub-regiões.

Os valores de dN e dS foram menores no grupo respondedor infectados pelo genótipo 1 e maiores no grupo respondedor infectado com o genótipo 3, comparado aos outros grupos testados. Não foi possível associar o valor da razão dN/dS ao resultado da terapia e também não foi observado padrão algum que pudesse ser associado ao sucesso ou insucesso terapêutico. Os tipos de resposta plotados nas topologias geradas a partir da região NS5A inteira nos dois genótipos virais também não foram agrupados, o que implica falta de associação entre a filogenia dos genótipos e a resposta ao tratamento. Apesar dos valores de dN/dS estatisticamente significantes encontrados na comparação entre os grupos não, foi possível detectar evidência de pressão seletiva positiva, visto que todos os valores da razão foram menores que 1, o que indica seleção purificadora em todos os grupos testados de pacientes infectados com vírus de genótipos 1 e 3.

A ausência de resultados estatisticamente significativos que pudessem ser correlacionados com a evidência de pressão seletiva positiva nesses estudos pode ser atribuída ao uso de metodologias que analisam as sequências par a par para a região inteira. O uso somente da razão dN/dS para detecção de seleção positiva através do cálculo de dN e dS entre duas sequências se mostra pouco eficiente por poder ter seu efeito diluído. Apesar da grande taxa de substituição de nucleotídeos do vírus da hepatite C, a maioria dos seus sítios apresenta mutações sinônimas enquanto apenas alguns sítios apresentam evidência de seleção positiva. Com isso, a média da razão sobre o total dos sítios de uma região ser raramente maior que 1 (Yang, 2007). A seleção natural atua em apenas alguns sítios específicos em um gene, sendo que a maioria dos códons possui uma razão de dN/dS <1, enquanto apenas alguns sítios apresentam dN/dS >1. Os métodos que analisam sequências par a par tiram a média da razão por todos os sítios, o que ocasiona um valor de dN/dS sempre menor que 1. A melhor abordagem para detectar seleção positiva é utilizar modelos que detectem seleção atuando em linhagens específicas na topologia ou sítios individuais, que permitem variação dos valores de dN/dS entre os sítios dos códons.

No trabalho realizado para determinar a influência do sistema imune (de Queiróz *et al.*, 2010), através dos epítomos CD8<sup>+</sup> específicos contra o HCV, no resultado da terapia com interferon e ribavirina nós identificamos 4 epítomos da região NS5A com diferenças significantes entre os grupos de pacientes, apesar do tamanho amostral reduzido (n=11 pacientes). Os epítomos VLSDFKTWL, SVASSASQLSA e ARGSPPSVASSASQLSAPS mostram frequências maiores no grupo que responde comparado ao grupo não respondedor. Esse resultado indica que esses epítomos, especialmente o VLSDFKTWL, induzidos pelo interferon, promovem uma resposta imune eficaz contra o HCV influenciando no resultado do tratamento. Apesar de o epítomo LTDPSHITA ter apresentado maior variação de aminoácidos no grupo respondedor, essa variação pode ser explicada como uma resposta insuficiente contra esse epítomo no grupo não respondedor, visto que esse epítomo não promove a eliminação da infecção viral nos pacientes.

É importante salientar que o epítomo VLSDFKTWL apresentou 100% de frequência no grupo que responde à terapia e somente 48% no grupo que não responde. Porém, o grupo não respondedor tem dois perfis distintos de falha do tratamento de acordo com a abordagem de 3 grupos: Pacientes que não responderes completos, que são definidos pela contínua presença do RNA viral durante todo o tratamento e após 6 meses do fim da terapia e pacientes que apresentaram RNA viral negativo durante e no fim do tratamento, porém apresentaram um "rebote", ou seja, quando é detectado RNA viral após o fim do tratamento. Esses pacientes podem ser classificados como não respondedor de final de tratamento. Todos os pacientes não responderes completos não apresentaram vírus com esse epítomo e apenas um paciente não respondedor no fim do tratamento não apresentou vírus com esse epítomo. Este resultado sugere o papel importante do epítomo VLSDFKTWL no controle da viremia e na eliminação da infecção viral.

Em estudos longitudinais poderão ser identificadas mutações na região desse epítomo, induzidas por pressão positiva do sistema imune que promovem o escape viral, que seria um dos fatores responsáveis pela falha da terapia, como observado no vírus da Influenza H3. Foi demonstrado que a evolução do gene da hemaglutinina promove escape viral da resposta humoral devido a pressão positiva exercida pelo sistema imune que provoca eventos de *drift* antigênico (Blackburne *et al.*, 2008). Em estudos com HIV, pacientes que mantinham a replicação viral em níveis basais e indetectável desenvolviam cargas virais elevadas após o aparecimento de mutações de escape em regiões sobre pressão positiva (Bailey *et al.*, 2007). Todos esses resultados sugerem que a seleção positiva é uma característica da evolução viral e que regiões sob a influência dessa seleção poderá sofrer mutações antes da eliminação do

vírus pelo sistema imune.

A análise para determinar evidência de seleção positiva nas sequências da região NS5A dos pacientes revelou que os três grupos apresentavam o mesmo valor da razão dN/dS (0,09) e valor similar quando analisados juntamente (grupo respondedor e não respondedor), sem diferença estatisticamente significativa. Todos os três valores indicaram seleção purificadora. Entretanto, como discutido anteriormente, somente o uso da razão dN/dS de uma região não é interessante para detecção de seleção positiva. Na análise em que utilizaram-se os modelos de seleção de máxima verossimilhança, o resultado não mostrou evidência de seleção positiva no grupo que responde ao tratamento. Este resultado, juntamente com o mapeamento dos epítomos, indica que as respostas do sistema imune, acrescido ao tratamento, foram o fator que promoveu a eliminação da infecção viral antes do aparecimento de mutações de escape que melhorem o *fitness* viral, o que levaria a falha do tratamento.

Quando analisamos o grupo que não responde e os dois grupos conjuntamente nós observamos evidência de pressão seletiva positiva, o que sugere que a falha do tratamento no grupo não respondedor é associada com a maior variação na região NS5A, apesar dos dois aminoácidos detectados pelo método *Bayes Empirical Bayes* como selecionados positivamente não estarem em regiões de epítomos. O resultado do mapeamento e a evidência de seleção positiva em regiões fora dos epítomos indicam que essas mutações mantêm a replicação viral na presença de drogas antivirais e compensam o *fitness* viral em um ambiente desfavorável.

Estudos mostram que a variabilidade na região da proteína NS5A é associada com resposta do tratamento com interferon, especialmente na região determinante da sensibilidade ao INF (ISDR), o domínio ligante da proteína quinase ativada por RNA (PKR) e a alça V3 (Enomoto *et al.*, 1996; Gale *et al.*, 1998; Nousbaum *et al.*, 2000). Entretanto, o papel dos epítomos no controle da replicação viral na terapia com interferon ainda não foi estabelecido. Nossos resultados confirmam essa característica da região NS5A, associando essa variabilidade com a influência da resposta mediada por epítomos CD8<sup>+</sup> específicos contra o HCV, mostrando que o epítopo VLSDFKTWL é associado com a eliminação da infecção durante a terapia, sugerindo que a eficiência da resposta imune por células T CD8<sup>+</sup> pode aumentar, sendo um dos fatores determinantes do sucesso terapêutico. É importante observar que esse epítopo é descrito na região de sinal de retenção citoplasmática (CRS) da proteína NS5A (Ubani *et al.*, 2001), sendo que essa região da proteína ainda não havia sido associada com resposta imunológica nem com efetividade terapêutica.

Mesmo com o tamanho amostral pequeno, nós encontramos evidências da importância dos epítomos de células CD8<sup>+</sup> específicos contra o HCV no controle da viremia, na persistência da infecção em pacientes crônicos e mutações de escape, o que está de acordo com diversos estudos (Timm *et al.*, 2004; Guglietta *et al.*, 2005; Tester *et al.*, 2005). Esses resultados revelam a grande influência dos linfócitos T CD8<sup>+</sup> também durante o período de tratamento, sendo modulado positivamente pelo interferon.

Mutações de escape podem limitar ou inibir a ligação com o HLA promovendo perda na apresentação antígeno ligado ao MHC, assim, nós analisamos as mutações que ocorreram nas regiões dos epítomos para determinar em quais posições ocorreram variação e comparar com informações do banco de dados SYFPEITHI (Rammensee *et al.*, 1999). Com exceção do epítomo SVASSASQLSA, que apresentou mutação de valina para leucina na segunda posição do epítomo, todas as outras mutações que ocorreram não foram associadas com regiões de aminoácidos de ancoramento ou auxiliador, de acordo com o banco de dados SYFPEITHI. Nesse epítomo, a mutação de valina para leucina corresponde a um aminoácido de ancoramento esperado. Sem variação no grupo respondedor, o epítomo VLSDFKTWL, apresentou no grupo não respondedor mutação na 3<sup>a</sup> posição do epítomo onde serina foi substituída por treonina em 37,8% e por alanina em 12,6% dos casos. Esse resultado reforça a importância do papel dos epítomos no controle da viremia em associação com a terapia, visto que esses aminoácidos não são classificados como resíduos preferenciais no banco de dados para o alelo A2, sugerindo que essas são mutações de escape. Essas mutações poderiam impedir que o epítomo fosse apresentado ligado ao HLA na superfície celular da célula infectada, sendo assim “invisível” ao sistema imune.

É importante salientar que, além da hipótese sugerida nesse trabalho, fatores do hospedeiro como restrição genética do sistema imune, diversidade do receptor da célula T ou a intensidade da resposta das células CD4<sup>+</sup> auxiliaadoras, podem contribuir com o escape viral observado no grupo não respondedor.

Para confirmar essa hipótese, nós decidimos focalizar as análises nas regiões correspondentes aos epítomos e analisar os valores da razão dN/dS, dN e dS para cada um dos epítomos e verificar se existem diferenças estatisticamente significantes. Para os 3 epítomos que mostraram frequências maiores no grupo que responde ao tratamento nós encontramos valores da razão dN/dS menor do que no grupo que não responde, sugerindo que o pressão seletiva positiva nos epítomos no último grupo ( $p < 0,05$ ). Apesar de o epítomo LTDPSHITA ter apresentado dN/dS maior no grupo que responde ao tratamento, este epítomo não promove resposta imune suficiente no grupo não respondedor, sendo uma resposta ineficaz, devido ao

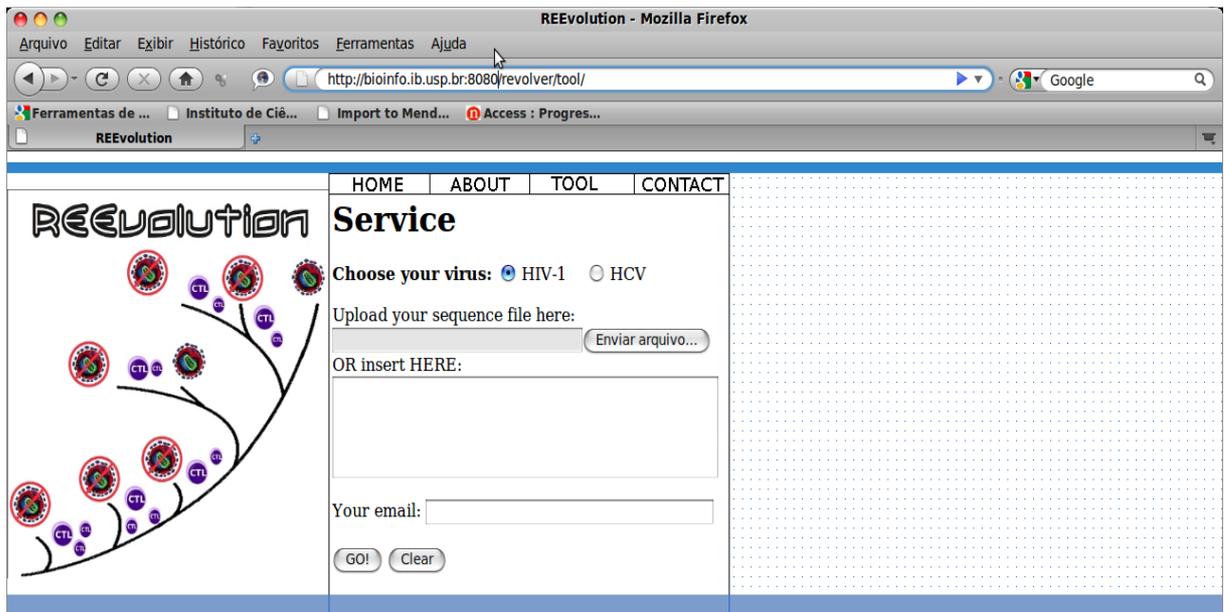
valor superior de dS no grupo que não responde e a falha no tratamento. Este resultado reforça a hipótese de resposta inespecífica a este epítopo.

Em resumo, o mapeamento dos 21 epítomos nas sequências dos 11 pacientes infectados pelo HCV do genótipo 1 do período pré-tratamento mostrou evidências de que ao menos 1 epítopo de célula T específico contra o HCV está associado com o resultado da terapia. Com os modelos de máxima verossimilhança confirmou-se a existência de seleção positiva no grupo não respondedor. Conjuntamente, esses resultados têm importantes implicações no entendimento da compreensão da dinâmica do sistema imune e o vírus durante o período do tratamento, confirmando o papel do sistema imune na eliminação da infecção.

No trabalho que realizamos com a proteína NS3, foi observada uma diferença entre as frequências dos epítomos entre os subtipos 1 e 3 (Carvalho-Mello *et al.*, 2009). Essa diferença entre os subtipos provavelmente deve-se a dois fatores: A grande variabilidade nucleotídica entre os grupos do HCV, que varia entre 30 e 35% entre os grupos e 20 e 25% entre os subgrupos (Simmonds *et al.*, 1994; Lau *et al.*, 1996), e a pouca representatividade de epítomos específicos do HCV genótipo 3 no banco de dados, devido ao fato do genótipo 1 ter uma prevalência mundial de 65% contra 30% do genótipo 3 (Mellor *et al.*, 1995; Codes *et al.*, 2003; Campiotto *et al.*, 2005). A grande variação entre os genótipos virais evidenciou a carência do banco de dados utilizados em epítomos específicos para o HCV genótipo 3, visto que esse subtipo tem uma prevalência relativamente pequena comparada com aquela do genótipo 1. O mapeamento dos epítomos nas sequências do genótipo 3 com a utilização desse banco de dados resultaria em uma subestimativa de epítomos desse genótipo com associação a resposta ao tratamento. Dessa forma, a análise de mapeamento dos epítomos não foi realizada nas sequências virais dos pacientes infectados por genótipo 3 por insuficiência de tamanho amostral adequado. Há que se destacar também que há um viés no banco de dados que favorece epítomos do genótipo 1. Assim, para a análise do mapeamento de epítomos, somente as sequências dos pacientes infectados pelo genótipo 1 foram selecionadas.

Para o desenvolvimento do programa *REEvolution*, outro banco de dados não curado foi selecionado como fonte dos epítomos (Vita *et al.*, 2010) permitindo o aumento de epítomos analisados pela ferramenta, auxiliando assim a sua acuidade. Além disso, diferentemente das ferramentas desenvolvidas para predição de epítomos, que analisam somente uma sequência de proteína por análise (Lundegaard *et al.*, 2007; Lafuente e Reche, 2009), nossa abordagem de análise de múltiplas sequências a partir de um alinhamento oferecido pelo usuário permite análises evolutivas que são indisponíveis em outras ferramentas de predição. Além disso, a

ferramenta utiliza epítomos descritos especificamente para o vírus analisado (HCV ou HIV), o que permite uma maior especificidade. A análise com modelos evolutivos, juntamente com o mapeamento dos epítomos, gera resultados que possibilitam a identificação de regiões que não estão sob influência do sistema imune, regiões reconhecidas pelo sistema imune, na identificação de mutações de escape na região do epítomo e identificação de mutações compensatórias que ocorram em regiões não reconhecidas pelo sistema imune.



**Figura 5** – Interface da ferramenta *REEvolution*.

Estas informações são úteis para o desenvolvimento de vacinas específicas contra esses vírus e para a melhoria de estratégias terapêuticas na hepatite C e na síndrome da imunodeficiência adquirida. A ferramenta *REEvolution* usa interface amigável e os resultados são de fácil compreensão, permitindo o usuário acesso à informação imunológica e evolutiva dos seus dados, como apresentado nas figuras 5 e 6. A ferramenta foi desenvolvida para ser visualizada em qualquer navegador de internet, porém, foi testada apenas nos navegadores *iExplorer*, *Mozilla Firefox* e *Safari*.

Entretanto, a análise do desempenho revelou que o tempo necessário para analisar  $N$  sequências aumenta quadraticamente devido aos modelos de evolução. Devido a esse fato, um limite no número de sequências será sugerido ao usuário, limitando à no máximo 50 sequências de 1500 nucleotídeos, ou um arquivo de 80 Kbytes.

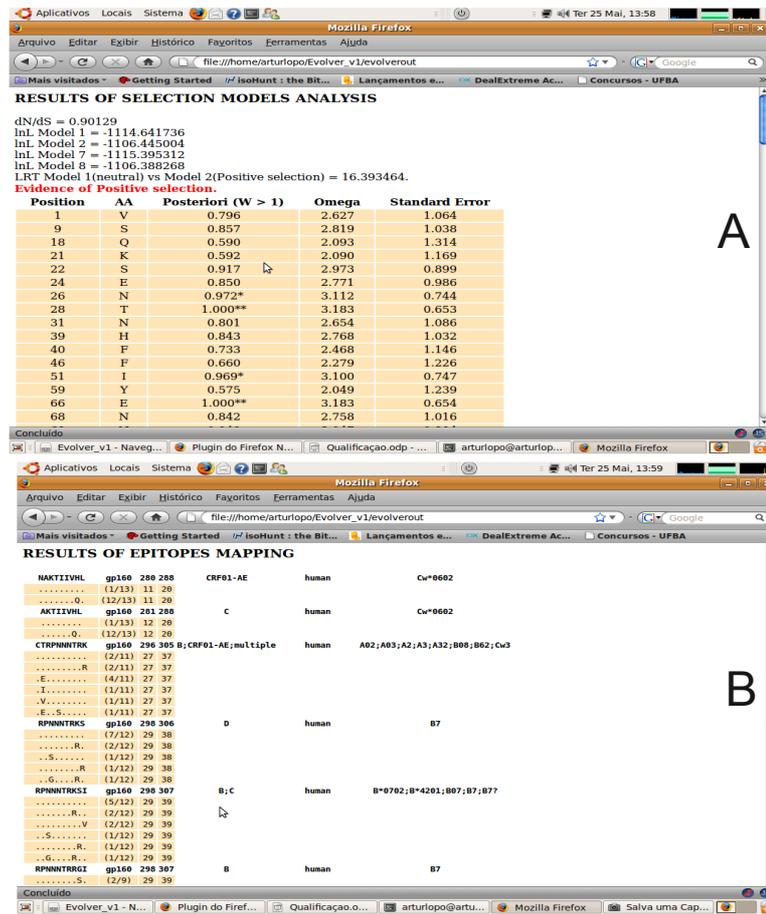


Figura 6 – Exemplo de resultados gerados pela ferramenta REEvolution. A – Resultado das análises de seleção. B -Resultado do mapeamento dos epítomos.

## Considerações Finais

## 7 – CONSIDERAÇÕES FINAIS

O resultado dos quatro artigos e o manuscrito submetido confirmam a importância do desenvolvimento da ferramenta para análise automática dos epítomos e da pressão seletiva para identificação de epítomos evolutivamente estáveis, fatores de predição do prognóstico e definição de grupos de risco. Para a versão final do programa utilizamos o banco de dados *immuneepitope*, que é mais abrangente, na qual mais informação adicionada auxiliará no aumento do poder de inferência das análises. A união da informação do mapeamento dos epítomos junto com a informação da história evolutiva, identificando eventos de pressão seletiva permitem uma associação mais clara e evidente dos processo de seleção impostos às cepas virais, sendo mais realístico em comparação a mapeamentos e predições de epítomos somente, visto que essa abordagem não permite acesso à informação evolutiva das amostras. Nossa ferramenta também permite a identificação de outros fatores de seleção, apenas com algumas adaptações no código, permitindo a implementação de mutações que conferem resistência à medicamentos antivirais, como observados na região NS3/4 do HCV em relação aos inibidores de polimerase, ou outros tipos de terapia.

Além das limitações orçamentárias, a disponibilidade de pacientes que se mantém nos programas de tratamento por longos períodos é bastante restrita. De qualquer forma, os resultados significativos obtidos nessas condições de baixo poder estatístico somente reforçam a importância dos resultados obtidos e a factibilidade da abordagem empregada. O desenvolvimento de ferramentas dedicadas poderá facilitar próximos estudos dessa natureza.

## Referências

## 8 – REFERÊNCIAS

- Bailey JR, Williams TM, Siliciano RF, Blankson JN. (2006). Maintenance of viral suppression in HIV-1-infected HLA-B\*57+ elite suppressors despite CTL escape mutations. *J Exp Med*. **15**;203(5):1357-69.
- Barnes E, Harcourt G, Brown D, Lucas M, Phillips R, Dusheiko G, Klenerman P. (2002). The dynamics of T lymphocyte responses during combination therapy for chronic hepatitis C virus infection. *Hepatology*. **36**: 743–54
- Bartenschlager R. (2002). Hepatitis C virus replicons: potential role for drug development. *Nat Rev Drug Discov*. **1**(11): 911-6.
- Bartosch B, Dubuisson J, Cosset FL. (2003). Infectious hepatitis C virus pseudo-particles containing functional E1-E2 envelope protein complexes. *J Exp Med*. **197**(5): 633-42.
- Bigger CB, Brasky KM, Lanford RE. (2001) DNA microarray analysis of chimpanzee liver during acute resolving hepatitis C virus infection. *J. Virol*. **75**:7059–66
- Bittar C, Jardim AC, Yamasaki LH, de Queiroz AT, Carareto CM, Pinho JR, de Carvalho-Mello IM, Rahal P. (2010). Genetic diversity of NS5A protein from hepatitis C virus genotype 3a and its relationship to therapy response. *BMC Infect Dis*. **23**:10(1):36. [Epub ahead of print]
- Blackburne BP, Hay AJ, Goldstein RA. (2008). Changing selective pressure during antigenic changes in human influenza H3. *PLoS Pathog*. **2**;4(5):e1000058.
- Bonhoeffer S, Holmes Ec, Nowak MA. (1995). Causes of HIV diversity. *Nature*. **376**:125.
- Bowen DG, Walker CM. (2005) Adaptive immune responses in acute and chronic hepatitis C virus infection. *Nature*. **436**:946–52
- Brumme ZL, . (2009) HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. *Plos One*. 4:8, e6687.
- Carvalho-Mello IM, Thumann C, Schvoerer E, Soulier E, Pinho JR, Silvestre DA, Queiroz AT, Wolf P, Baumert TF, Keller FS, Pereira CA. (2009). Conservation of hepatitis C virus nonstructural protein 3 amino acid sequence in viral isolates during liver transplantation. *J Viral Hepat*. **16**(10):732-7.
- Campiotto S, Pinho JR, Carrilho FJ, Da Silva LC, Souto FJ, Spinelli V, Pereira LM, Coelho HS, Silva AO, Fonseca JC, Rosa H, Lacet CM, Bernardini AP. (2005). Geographic distribution of hepatitis C virus genotypes in Brazil. *Braz J Med Biol Res*. **38**(1): 41-9.
- Codes L, de Freitas LA, Santos-Jesus R, Vivitski L, Silva LK, Trepo C, Reis MG, Parana R. (2003). Comparative study of hepatitis C virus genotypes 1 and 3 in Salvador, Bahia Brazil. *Braz J Infect Dis*. **7**(6): 409-17.
- Choo QL, Richman KH, Han JH, Berger K, Lee C, Dong C, Gallegos C, Coit D, Medina-Selby R, Barr PJ, Weiner AJ, Bradley DW, Kuo G, Houghton M. (1991) Genetic

- organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci U S A.* **88**(6): 2451-5.
- Cooper S, Erickson AL, Adams EJ, Kansopon J, Weiner AJ, et al. Analysis of a successful immune response against hepatitis C virus. *Immunity.* 1999; **10**:439–49
- Cox AL, Mosbruger T, Lauer GM, Pardoll D, Thomas DL, Ray SC. (2005). Comprehensive analyses of CD8+ Tcell responses during longitudinal study of acute human hepatitis C. *Hepatology.* **42**:104–12
- Cox AL, Mosbruger T, Mao Q, Liu Z, Wang XH, Yang HC, Sidney J, Sette A, Pardoll D, Thomas DL, Ray SC. (2005). Cellular immune selection with hepatitis C virus persistence in humans. *J Exp Med.* **201**(11):1741–1752.
- Dayhoff MO, Schwartz RM, Orcutt BC. (1978). A model of evolutionary change in proteins. *Atl Prot Seq and Struc.* **5**(3): 345–352.
- Deleersnyder V, Pillez A, Wychowski C, Blight K, Xu J, Hahn YS, Rice CM, Dubuisson J. (1997). Formation of native hepatitis C virus glycoprotein complexes. *J Virol.* **71**(1): 697-704.
- de Queiróz ATL, Maracajá-Coutinho V, Jardim ACG, Rahal P, Carvalho-Mello IMVG, Matioli SR. (2010) Relation of Pretreatment Sequence Diversity in NS5A Region of HCV genotype 1 With Immune Response Between pegylated-IFN/ribavirin Therapy Outcomes. *J Viral Hept.* [Epub ahead of print] <http://www.ncbi.nlm.nih.gov/pubmed/20456637>
- Domingo E, Holland JJ. (1997) RNA virus mutations and fitness for survival. *Annu Rev Microbiol.* **51**:151-178
- Drucker E, Alcabes PG, Marx PA. (2001). The injection century: massive unsterile injections and the emergence of human pathogens. *Lancet.* **358**(9297):1989-92
- Duffy S, Shackelton LA, Holmes EC. (2008). Rates of Evolutionary change in viruses: patterns and determinants. *Nat Rev.* **9**: 267-75.
- Eddy SR. (1996). Hidden Markov models. *Curr Opin Struct Biol.* **6**:361–5.
- Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nuc Acid Res.* **35**(5):1792-97.
- Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Ogura Y, Izumi N, Marumo F, Sato C. (1996). Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N Engl J Med.* **334**(2):77-81.
- Erickson AL, Kimura Y, Igarashi S, Eichelberger J, Houghton M, Sidney J, McKinney D, Sette A, Hughes AL, Walker CM. (2001). The outcome of hepatitis C virus infection is predicted by escape mutations in epitopes targeted by cytotoxic T lymphocytes. *Immunity.* **15**:883–95

- Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJ, Hofmann K, Bairoch A. (2002). The Prosite database, its status in 2002. *Nucleic Acids Resource*. **30**:235-238.
- Feld JJ, Hoofnagle JH. (2005). Mechanism of action of interferon and ribavirin in treatment of hepatitis C. *Nature*. **436**(7053):967-72.
- Flajnik MF, du Pasquier L. (2004). Evolution of innate and adaptative immunity: can we draw a line? *Trend Immunol*. **25**:640-44.
- Gale MJ, Korth MJ, Katze MG. (1998). Repression of the PKR protein kinase by the hepatitis C virus NS5A protein: a potential mechanism of interferon resistance. *Clin Diagn Virol*. **10**(2-3):157-162.
- Gerrish PJ, Garcia-Laerma JG. (2003). Mutation rate and the efficacy of antimicrobial drug treatment. *Lancet Infect Dis*. **3**:28-32.
- Gimenez-Barcons M, Wang C, Chen M, Sanchez-Tapias JM, Saiz JC, Gale MJ. (2005). The oncogenic potential of hepatitis C virus NS5A sequence variants is associated with PKR regulation. *J Interferon Cytokine Res*. **25**:152-64.
- Goldman N, Yang Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*. **11**(5):725-36.
- Goulder PJ, Watkins DI. (2008). Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nat Rev Immunol*. **8**:619-630.
- Grakoui A, Shoukry NH, Woollard DJ, Han JH, Hanson HL, Ghayeb J, Murthy KK, Rice CM, Walker CM. (2003). HCV persistence and immune evasion in the absence of memory T cell help. *Science*. **302**:659-62
- Guglietta S, Garbuglia AR, Pacciani V, Scottà C, Perrone MP, Laurenti L, Spada E, Mele A, Capobianchi MR, Taliani G, Folgori A, Vitelli A, Ruggeri L, Nicosia A, Piccolella E, Del Porto P. (2005). Positive selection of cytotoxic T lymphocyte escape variants during acute hepatitis C virus infection. *Eur J Immunol*. **35**(9):2627-2637.
- Hall TA. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser*. **41**:95-98.
- Hardin C, Pogorelov TV, Luthey-Schulten Z. (2002). Ab initio protein structure prediction. *Curr Opin Struct Biol*. **12**:176-81.
- Henikoff S, Henikoff JG. (1992). Amino Acid Substitution Matrices from Protein Blocks. *Proc Nat Acad Soc*. **89**(22): 10915-10919.
- Hijikata M, Mizushima H, Akagi T, Mori S, Kakiuchi N, Kato N, Tanaka T, Kimura K, Shimotohno K. (1996). Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. *J Virol*. **67**(8): 4665-75.
- James A, 2001. Ilustração do vírus da hepatite C. Disponível em:

[www.rit.edu/~japfaa/hcv.jpg](http://www.rit.edu/~japfaa/hcv.jpg).

- Jardim AC, Yamasaki LH, de Queiróz AT, Bittar C, Pinho JR, Carareto CM, Rahal P, Mello IM. (2009). Quasispecies of hepatitis C virus genotype 1 and treatment outcome with Peginterferon and Ribavirin. *Infect Gen Evol.* **9**(4):689-98.
- Jardetzky TS, Brown JH, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC. (1996) Crystallographic analysis of endogenous peptides associated with HLADR1 suggests a common, polyproline II-like conformation for bound peptides. *Proc Natl Acad Sci U S A.* **93**:734–738.
- Jenkins GM, Rambaut A, Pybus OG, Holmes EC. (2002) Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J mol Evol.* **54**:156-165
- Johnson CL, Gale MJ. (2006). CARD games between virus and host get a new player. *Trends Immunol.* **27**:1–4.
- Kamal SM, Fehr J, Roesler B, Peters T, Rasenack JW. (2002). Peginterferon alone or with ribavirin enhances HCV-specific CD4 T-helper 1 responses in patients with chronic hepatitis C. *Gastroenterology.* **123**:1070–83
- Katze MG, He Y, Gale M, Jr. (2002). Viruses and interferon: a fight for supremacy. *Nat Rev Immunol.* **2**(9): 675-87.
- Koonin EV, Senkevich TG, Dolja VV. (2006). The ancient Virus World and evolution of cells. *Biol Direct.* **19**:1:29.
- Korber B. (2000). HIV Signature and Sequence Variation Analysis. Computational Analysis of HIV Molecular Sequences, Chapter 4, pages 55-72. Ed. Allen G. Rodrigo and Gerald H. Learn. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Korber BT. (2007). HIV Molecular Immunology. Ed. Korber BTM, Brander C, Haynes BF, Koup R, Moore JP, Walker BD, Watkins DW. Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico.
- Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. (2007). Large scale validation methods for cytotoxic T lymphocyte epitope prediction. *BMC Bioinformatics.* **31**(8):402
- Lau JY, Davis GL, Prescott LE, Maertens G, Lindsay KL, Qian K, Mizokami M, Simmonds P. (1996). Distribution of hepatitis C virus genotypes determined by line probe assay in patients with chronic hepatitis C seen at tertiary referral centers in the United States. Hepatitis Interventional Therapy Group. *Ann Intern Med.* **124**(10): 868-76.
- Lechner F, Wong DK, Dunbar PR, Chapman R, Chung RT, Dohrenwend P, Robbins G, Phillips R, Klenerman P, Walker BD. (2000). Analysis of successful immune responses in persons infected with hepatitis C virus. *J Exp Med.* **191**:1499–512
- Li, WH. (1993). Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* **36**:96–99.

- Lin C, Lindenbach BD, Pragai BM, McCourt DW, Rice CM. (1994). Processing in the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini. *J Virol.* **68**(8): 5063-73.
- Lundegaard C, Lund O, Kesmir C, Brunak S, Nielsen M. (2007). Modeling the adaptive immune system: predictions and simulations. *Bioinformatics.* **23**(24) 3265-75.
- Lafuente E.M., Reche P.A. (2009) Prediction of MHC-peptide binding: a systematic and comprehensive overview. *Curr Pharm Des.* **15**(28) 3209-20.
- Maenaka K, Jones EY. (1999). MHC superfamily structure and the immune system. *Curr Opin Struct Biol.* **9**: 745-53.
- Magiorkinis G, Magiorkinis E, et al. The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. *PLoS Med.* 2009; **6**(12):e1000198.
- Markov PV, Pepin J, Frost E, Deslandes S, Labbé AC, Pybus OG. (2009). Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. *J Gen Virol.* **90**(9):2086-96.
- Martell M, Esteban JI, Quer J, Genescà J, Weiner A, Esteban R, Guardia J, Gómez J. (1992). Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J Virol.* **66**(5): 3225-9.
- McHutchison JG & Poynard T. (1999). Combination therapy with interferon plus ribavirin for the initial treatment of chronic hepatitis C. *Semin Liver Dis.* **19** 1:, 57–65.
- Mellor J, Holmes EC, Jarvis LM, Yap PL, Simmonds P. (1995). Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: implications for virus classification. The International HCV Collaborative Study Group. *J Gen Virol.* **76** (10): 2493-507.
- Miller RH, Purcell RH. (1990). Hepatitis C virus share amino acid sequence similarity with pestiviruses and flaviviruses as well as member of two plant virus supergroup. *Proc Nat Acad Sci.* **87**: 2057-2061.
- Mizokami M, Gojobori T, Ohba K, Ikeo K, Ge XM, Ohno T, Orito E, Lau JY. (1996) Hepatitis C virus types 7, 8 and 9 should be classified as type 6 subtypes. *J Hepatol.* **24**(5):622-4.
- Miyata T and Yasunaga T. (1980). Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its applications. *J. Mol. Evol.* **16**:23–36.
- Nakano T, Lu L, Liu P, Pybus OG. (2004). Viral gene sequences reveal the variable history of hepatitis C virus infection among countries. *J Infect Dis.* **190**(6):1098-108.
- Nei M and Gojobori T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418–426.

- Neumann-Haefelin C, McKiernan S, Ward S, Viazov S, Spangenberg HC, Killinger T, Baumert TF, Nazarova N, Sheridan I, Pybus O, von Weizsäcker F, Roggendorf M, Kelleher D, Klenerman P, Blum HE, Thimme R. (2006). Dominant influence of an HLA-B27 restricted CD8<sup>+</sup> T cell response in mediating HCV clearance and evolution. *Hepatology*. **43**:563–72
- Nicholas KB, Nicholas HBJ, Deerfield DW. (1997). GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNEW NEWS*. **4**:1-4.
- Nielsen M, Lundegaard C, Lund O. 2007. Prediction of MHC class II binding affinity using SMMalign, a novel stabilization matrix alignment method. *BMC Bioinformatics* **8**: 238.
- Nousbaum J, Polyak SJ, Ray SC, Sullivan DG, Larson AM, Carithers RL Jr, Gretch DR. (2000). Prospective characterization of full-length hepatitis C virus NS5A quasispecies during induction and combination antiviral therapy. *J Virol* **74**(19):9028-9038.
- Orr HA.(2009). Fitness and its role in evolutionary genetics. *Nat Rev Genet*.**10**(8):531-9.
- Pancholi P, Perkus M, TricocheN, Liu Q, Prince AM. (2003). DNA immunization with hepatitis C virus (HCV) polycistronic genes or immunization by HCV DNA priming-recombinant canarypox virus boosting induces immune responses and protection from recombinant HCV-vaccinia virus infection in HLA-A2.1-transgenic mice. *J Virol*. **77**(1): 382–390
- Pearson WR, Lippman DJ. (1988). Improved Tools for Biological Sequence Comparison. *Proc Natl Acad Sci U S A*. **85**:2444-2448
- Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, Fleri W, Kronenberg M, Kubo R, Lund O, Nemazee D, Ponomarenko JV, Sathiamurthy M, Schoenberger S, Stewart S, Surko P, Way S, Wilson S, Sette A.(2005). The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol* **3**(3):e91.
- Pileri P, Uematsu Y, Campagnoli S, Galli G, Falugi F, Petracca R, Weiner AJ, Houghton M, Rosa D, Grandi G, Abrignani S. (1998). Binding of hepatitis C virus to CD81. *Science*. **282**(5390): 938-41.
- Plamondon M, Labbé AC, Frost E, Deslandes S, Alves AC, Bastien N, Pepin J. (2007). Hepatitis C virus infection in Guinea-Bissau: a sexually transmitted genotype 2 with parenteral amplification?. *PLoS One*. **2**(4):e372.
- Racanelli V, Rehmann B. (2003). Hepatitis C virus infection: when silence is deception. *Trends Immunol*. **24**:456–64.
- Rammensee HG, Friede T, Stevanović S. (1995). MHC Ligands and peptide motifs: First listing. *Immunogenetics*. **41**:178–228.
- Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanović S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50**(3-4):213-9.

- Roberts SK, Cooksley G, Dore GJ, Robson R, Shaw D, Berns H, Hill G, Klumpp K, Najera I, Washington C. (2008). Robust antiviral activity of R1626, a novel nucleoside analog: a randomized, placebo-controlled study in patients with chronic hepatitis C. *Hepatology*. **48**(2):398-406.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*. **19**: 2496–2497.
- Schroter M, Zollner B, Laufs R, Feucht HH. (2004). Changes in the prevalence of hepatitis C virus genotype among injection drug users: A highly dynamic process. *J Infect Dis*. **190**:1199–1200.
- Shoukry NH, Grakoui A, Houghton M, Chien DY, Ghayeb J, Reimann KA, Walker CM. (2003). Memory CD8+ T cells are required for protection from persistent hepatitis C virus infection. *J Exp Med*. **197**:1645–55
- Shukla DD, Hoyne PA, Ward CW. (1995). Evaluation of complete genome sequences and sequences of individual gene products for the classification of hepatitis C viruses. *Arch Virol*. **140**(10): 1747-61.
- Silverstein AM. (2003). Cellular versus humoral immunity:a century long dispute. *Nat immunol*. **4**: 425-8.
- Simmonds P, Alberti A, Alter HJ, Bonino F, Bradley DW, Brechot C, Brouwer JT, Chan SW, Chayama K, Chen DS, et al. (1994). A proposed system for the nomenclature of hepatitis C viral genotypes. *Hepatology*. **19**(5): 1321-4.
- Simmonds P. (2001). The origin and evolution of hepatitis viruses in humans. *J Gen Virol*. **82**: 693–712.
- Simmonds P. (2004). Genetic diversity and evolution of hepatitis C virus--15 years on. *J Gen Virol*. **85**(11): 3173-88.
- Simmonds P, Bukh J, Combet C, Deléage G, Enomoto N, Feinstone S, Halfon P, Inchauspé G, Kuiken C, Maertens G, Mizokami M, Murphy DG, Okamoto H, Pawlotsky JM, Penin F, Sablon E, Shin-I T, Stuyver LJ, Thiel HJ, Viazov S, Weiner AJ, Widell A. (2005). Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology*. **42**: 962–973
- Smith DB, Pathirana S, Davidson F, Lawlor E, Power J, Yap PL, Simmonds P. (1997). The origin of hepatitis C virus genotypes. *J Gen Virol*. **78** (2):321-8.
- Stimson GV. (1996). The global diffusion of injecting drug use: implications for human immunodeficiency virus infection. *Bull Narc*. **45**: 3–17.
- Suzuki Y, Gojobori T. (1999). A method for detecting positive selection at single amino acid site. *Mol Biol Evol*. **16**:1315-1328.
- Tamura K, Dudley J, Nei M, Kumar S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol Biol Evol*. **24**, 1596–1599.

- Tanaka Y, Hanada K, Mizokami M, Yeo AE, Shih JW, Gojobori T, Alter HJ. (2002). Inaugural article: a comparison of the molecular clock of hepatitis C virus in the United States and Japan predicts that hepatocellular carcinoma incidence in the United States will increase over the next two decades. *Proc Natl Acad Sci U S A*. **99**:15584–15589.
- Taylor DR, Shi ST, Romano PR, Barber GN, Lai MMC. (1999). Inhibition of the interferon-inducible protein kinase PKR by the HCV E2 protein. *Science*. **285**:107–10
- Tester I, Smyk-Pearson S, Wang P, Wertheimer A, Yao E, Lewinsohn DM, Tavis JE, Rosen HR. (2005). Immune evasion versus recovery after acute hepatitis C virus infection from a shared source. *J Exp Med* **201**(11):1725-1731.
- Terasaki PI. (2007). A brief history of HLA. *Immunol Res*. **38**: 139-48.
- Thompson A, Patel K, Tillman H, McHutchison JG. (2009). Directly acting antivirals for the treatment of patients with hepatitis C infection: a clinical development update addressing key future challenges. *J. Hepatol*. **50**:184–194.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997). The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. **25**(24):4876-4882.
- Thimme R, Lohmann V, Weber F. (2006). A target on the move: innate and adaptive immune escape strategies of hepatitis C virus. *Antiviral Res*. **69**:129–41
- Timm J, Lauer GM, Kavanagh DG, Sheridan I, Kim AY, Lucas M, Pillay T, Ouchi K, Reyor LL, Schulze zur Wiesch J, Gandhi RT, Chung RT, Bhardwaj N, Klenerman P, Walker BD, Allen TM. (2004). CD8 epitope escape and reversion in acute HCV infection. *J Exp Med*. **200**:1593–604.
- Tong X, Bogen S, Chase R, Girijavallabhan V, Guo Z, Njoroge FG, Prongay A, Saksena A, Skelton A, Xia E, Ralston R. (2008). Characterization of resistance mutations against HCV ketoamide protease inhibitors. *Antiviral Res*. **77**(3):177-85.
- Ubani S, Uggeri J, Matsuura Y, Miyamura T, Penna A, Boni C, Ferrari C. (2001). Identification of immunodominant hepatitis C virus (HCV)-specific cytotoxic T-cell epitopes by stimulation with endogenously synthesized HCV antigens. *Hepatology*. **33**(6):1533-43.
- van Asten L, Verhaest I, Lamzira S, Hernandez-Aguado I, Zangerle R, Boufassa F, Rezza G, Broers B, Robertson JR, Brettler RP, McMenemy J, Prins M, Cochrane A, Simmonds P, Coutinho RA, Bruisten S et al. (2004). Spread of hepatitis C virus among European injection drug users infected with HIV: A phylogenetic analysis. *J Infect Dis*. **189**:292–302.
- Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A, Peters B. (2010). The immune epitope database 2.0. *Nucleic Acids Res*. **38**, D854-62.
- Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. (2008). A systematic assessment of

MHC class LL peptide binding predictions and evaluation of a consensus approach. *PLoS Comput Biol.* 4(4):e1000048.

- Wieland SF, Chisari FV. (2005). Stealth and cunning: hepatitis B and hepatitis C viruses. *J Virol.* 79:9369–80
- World Health Organization. (1999). Global surveillance and control of hepatitis C. Report of WHO consultation organized in collaboration with the Viral Hepatitis Prevention Board, Antwerp, Belgium. *J Viral Hepat.* 6:35-47.
- Wohnslund A, Hofmann WP, Sarrazin C. (2007). Viral determinants of resistance to treatment in patients with hepatitis C. *Clin Microbiol Rev.* 20(1):23-38.
- Yang, Z., and R. Nielsen. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution.* 17:32-43.
- Yang Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Yang Z, Wong WSW, Nielsen R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 22:1107–1118.
- Yi S, Ellsworth DL, Li WH. (2002). Slow molecular clocks in Old World monkeys, apes, and humans. *Mol Biol Evol.* 19(12):2191-8.
- Yusim K, Richardson R, Tao N, Dalwani A, Agrawal A, Szinger J, Funkhouser R, Korber B, Kuiken C. (2005). Los alamos hepatitis C immunology database. *Appl Bioinformatics.* 4(4):217-25.
- Zeuzem S, Schmidt JM, Lee JH, Rüster B, Roth WK. (1996). Effect of interferon alfa on the dynamics of hepatitis C virus turnover in vivo. *Hepatology.* 23(2): 366-71.
- Zhou S, Terrault NA, Ferrell L, Hahn JA, Lau JY, Simmonds P, Roberts JP, Lake JR, Ascher NL, Wright TL. (1996). Severity of liver disease in liver transplantation recipients with hepatitis C virus infection: relationship to genotype and level of viremia. *Hepatology.* 24(5): 1041-6.

**ANEXOS**

## 9 – ANEXOS

1 – Artigo: Jardim AC, Yamasaki LH, de Queiróz AT, Bittar C, Pinho JR, Carareto CM, Rahal P, Mello IM. Quasispecies of hepatitis C virus genotype 1 and treatment outcome with Peginterferon and Ribavirin. 2009. *Infectious Genetics and Evolution*.

2 – Artigo: Mello IM, Thumann C, Schvoerer E, Soulier E, Pinho JR, Silvestre DA, Queiroz AT, Wolf P, Baumert TF, Keller FS, Pereira CA. Conservation of hepatitis C virus nonstructural protein 3 amino acid sequence in viral isolates during liver transplantation. 2009. *Journal of Viral Hepatitis*.

3 – Artigo: Bittar C, Jardim AC, Yamasaki LH, de Queiroz AT, Carareto CM, Pinho JR, de Carvalho-Mello IM, Rahal P. Genetic diversity of NS5A protein from hepatitis C virus genotype 3a and its relationship to therapy response. 2010. *BMC Infectious Diseases*.

4 – Artigo: de Queiróz ATL, Maracajá-Coutinho V, Jardim ACG, Rahal P, Carvalho-Mello IMVG, Matioli SR. Relation of Pretreatment Sequence Diversity in NS5A Region of HCV genotype 1 With Immune Response Between pegylated-INF/ribavirin Therapy Outcomes. 2010, *Journal of Viral Hepatitis*.

5 – Manuscrito Submetido: de Queiróz ATL, Rangel LT, Matioli SR, Carvalho-Mello IM. REEvolution: A tool for epitope mapping and selective pressure analysis for HIV and HCV infections.