unesp
UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José do Rio Preto

Lílian Hiromi Tomonari Yamasaki

Pirossequenciamento de alta cobertura da região hipervariável 1 do Vírus da Hepatite C

São José do Rio Preto – SP
Abril/2014

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José do Rio Preto

Lílian Hiromi Tomonari Yamasaki

Pirossequenciamento de alta cobertura da região hipervariável 1 do Vírus da Hepatite C

Tese apresentada como parte das atividades para a obtenção do título de Doutor do Programa de Pós-Graduação em Microbiologia do Instituto de Biociências, Letras e Ciências Exatas – Universidade Estadual Paulista "Júlio de Mesquita Filho"

Área de concentração: Virologia

Orientadora: Profª.Drª. Paula Rahal (UNESP/IBILCE)

Co-orientador: Dr. Yury Khudyakov (CDC/DVH-Atlanta/USA)

São José do Rio Preto – SP
Abril/2014

Yamasaki, Lílian Hiromi Tomonari.
Pirossequenciamento de alta cobertura da região hipervariável 1 do vírus da hepatite C / Lílian Hiromi Tomonari Yamasaki. -- São José do Rio Preto, 2014
65 f. : il., gráfs.

Orientador: Paula Rahal
Coorientador: Yury Khudyakov
Tese (doutorado) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas

1. Virologia. 2. Hepatite C. 3. Vírus de RNA. 4. Vírus - Aspectos genéticos. 5. Agentes antivirais. 6. Interferon. I. Rahal, Paula. II. Khudyakov, Yury. III. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências Exatas. IV. Título.

CDU – 576.858

Ficha catalográfica elaborada pela Biblioteca do IBILCE
UNESP - Câmpus de São José do Rio Preto

Autora: Lílian Hiromi Tomonari Yamasaki

Título: Pirossequenciamento de alta cobertura da região hipervariável 1 do Vírus da Hepatite C

Comissão Examinadora

Prof(a). Dr(a). Ana Carolina Gomes Jardim
Presidente
Depto. de Área de Imunologia Microbiologia e Parasitologia Universidade Federal de Uberlândia (UFU) – MG

Prof(a). Dr(a). Isabel Maria Vicente Guedes de Carvalho Mello
Membro
Depto. de Medicina
Universidade Federal de São Paulo (UNIFESP) – SP

Dra. Camila Malta Romano
Membro
Instituto de Medicina Tropical de São Paulo – SP

Dr(a). Paola Jocelan Scarin Provazzi Trabulsi
Membro
Instituto de Biociencias, Letras e Ciencias Exatas (IBILCE)
Universidade Estadual Paulista “Julio de Mesquita Filho” (UNESP/IBILCE)

Dr(a). Cintia Bittar
Membro
Instituto de Biociencias, Letras e Ciencias Exatas (IBILCE)
Universidade Estadual Paulista “Julio de Mesquita Filho” (UNESP/IBILCE)

Data da defesa: 30 de abril de 2014

Agradecimentos

- À coordenadora deste projeto Profa. Dra. Paula Rahal pela amizade, ensinamentos e a oportunidade de realizar este trabalho;
- Ao Dr. Yury Khudyakov e toda a sua equipe do CDC (Gilberto, Zoya, Lilia, Pavel, David, Forbi, Hong, James, Guo-Liang, Yulin, LaShonda, Paula, Livia, Aufra) pela colaboração e por me acolher nos Estados Unidos;
- Ao Lucas pela ajuda na organização dos dados e das amostras utilizadas neste trabalho;
- A Lenira Bueno por todo o auxílio na burocracia que faz parte da formação acadêmica e do laboratório;
- A todos os colegas do Genoma, pelos quase onze anos de companheirismo, amizade e risadas;
- A ex-co-orientadora Profa. Dra. Ana Carolina Gomes Jardim, por aceitar assumir a cadeira de presidente da banca examinadora;
- A banca examinadora, pelo tempo dedicado a melhorar este trabalho;
- Aos amigos presentes (e aos ausentes) e as moradoras da Casa Bonita (Ana, Monica e Nádia);
- A minha família em especial minha irmã Rosana pela ajuda com as análises (mesmo sem entender a finalidade destes);
- Ao apoio financeiro da Capes, FAPESP e do CDC.

SUMÁRIO

Capítulo 1

1. Introdução e Revisão Bibliográfica..02

1.1. A Hepatite C 02

1.2. O vírus da Hepatite C 04

1.3. A variabilidade do genoma viral e quasispecies 05

1.4. A região hipervariável 1 da proteína do envelope 2..07

1.5. Tratamento e tipos de resposta ao tratamento 09

1.6. Fatores preditivos de resposta ao tratamento baseado em IFN 11

1.7. Sequenciamento de nova geração aplicado a virologia 13

2. Justificativa 16

3. Objetivos..18

3.1. Objetivo geral..19

3.2. Objetivos específicos 19

4. Referencias Bibliográficas..20

Capítulo 2

1. Artigo Cientifico 1..26

2. Artio Cienífico 2..51

Capítulo 3

Conclusões gerais..64

Pirossequenciamento de alta cobertura da região hipervariável 1 do Vírus da Hepatite C

RESUMO

Embora novos medicamentos de ação direta anti-HCV tenham sido recentemente aprovados no mercado, com exceção do genótipo 1, a terapia mais utilizada ainda é baseada em Interferon (IFN) e Ribavirina. Desta forma, o genótipo 3 é o que apresenta mais baixa taxa de sucesso no tratamento, no atual cenário. Um dos fatores determinantes do insucesso deste tratamento no paciente é a variabilidade viral. O HCV apresenta alta taxa de mutação durante a replicação, implicando na produção de variantes intra-hospedeiro, denominadas *quasispecies*. A região hipervariável 1 (HVR1) da proteína do envelope é uma região de alta variabilidade do genoma viral. A detecção das quasispecies minoritárias pode ser realizada de forma confiável e eficiente por pirossequenciamento de alta cobertura (UPDS), técnica capaz de detectar variantes presentes em frequência <1% na população. Com base neste contexto, foram determinados quasispecies da HVR1-HCV de 14 pacientes infectados com o genótipo 3 do vírus, utilizando a técnica de UPDS. No total, 64.400 sequencias da HVR1 foram obtidas de amostras pré-tratamento. Destas, 27.398 sequências de alta qualidade foram filtradas e corrigidas. Valores de distância genética e entropia de Shannon não foram correlacionados a resposta ao tratamento. Estas sequências foram posteriormente submetidas a construção de redes median-joining e análise Bayesiana de estrutura populacional (BAPS). A análise identificou amostras com diferentes estruturas, desde altamente conservadas (apenas uma população de quasispecies) até altamente estratificadas (6 subpopulações). Este resultado foi condizente com a estrutura das redes *median-joining*. Várias mutações ao longo da HVR1 do mesmo paciente e algumas repetiram em pacientes do mesmo grupo de resposta. Quanto a avaliação de análise das sequências de aminoácidos, embora haja grande variação quanto a sequência e propriedades físico-químicas, pode-se observar conservação quanto a estrutura secundária, principalmente quanto a distribuição de voltas e *coils*. As regiões de epitopos se concentram na região carboxiterminal, e pode haver algumas variações entre as quasispecies quanto ao número e tamanho, mesmo dentro do mesmo paciente. Estes dados irão contribuir com o conhecimento sobre a composição das quasispecies de HCV e a terapia e como uma ferramenta de alta resolução como o UPDS é essencial nestes estudos.

Palavras-chave: Hepatite C, HCV-3, Interferon, quasispecies, pirossequenciamento

Hepatitis C Virus Quasispecies Analysis Using Ultra-Deep Pyrosequencing

ABSTRACT

Hepatitis C is a major public health problem. New HCV antiviral drugs were released on market on 2010; however, excluding for genotype 1, the most used therapy used currently is still based on Interferon (IFN) and Ribavirin. Nowadays, genotype 3 is the one with the highest rate of treatment failure. Viral genome variability is one of the factors that lead in therapy failure. HCV presents a high mutability during replication course, implicating in arising of intra-host variants called quasispecies. The hypervariable region 1 (HVR1) from envelope protein presents as quasispecies and may be related to IFN therapy resistance. Resistant quasispecies may not represent majority of variants population in the host, therefore, in these cases traditional sequencing techniques are unable to detect. For detection of minority quasispecies, ultra-deep pyrosequencing (UPDS) is a reliable and efficient tool, being able to detect even variants with frequency <1% in the population. Regarding this issue, we determined HVR1 quasispecies from 14 patients infected with HCV genotype 3 using the UPDS approach. In total, 64,400 HVR1 sequences were obtained from pre-therapy sample. From these sequences, 27,398 ones with high quality were filtered. Genetic distance and Shannon entropy values were not related to therapy outcome. These sequences were analyzed using median-joining networks and Bayesian population structural analysis. These analysis identified samples with different structures, from high conserved (one sub-population) to high stratified ones (6 sub-populations). Networks analysis also confirmed this result. Mutations exclusive for a type of response of therapy were identified along HVR1. Amino acid sequences indicated that this region presents conserved structure, even if sequence and physical and chemicals properties seem flexible. Especially in turns and coils positions, this conservation seems notable. Potential epitopes positions are concentrated in carboxiterminal and can vary of number and size between quasispecies These results will contribute to the understanding of HCV quasispecies dynamics and therapy and how a high resolution tool as UPDS is essential to it.

Keywords: Hepatitis C, HCV-3, Interferon, quasispecies, pirosequencing

Capítulo 1

Revisão Bibliográfica, Justificativa e Objetivos

1. Introdução e Revisão Bibliográfica

1.1. A Hepatite C

A Hepatite C é considerada a principal hepatite não-A e não-B associada às transfusões sanguíneas e adquirida na comunidade. A doença pode evoluir para distúrbios malignos crônicos, incluindo cirrose hepática e carcinoma hepatocelular (ALTER et al., 1989; SAITO et al., 1990). A alta probabilidade de persistência da doença é característica da infecção pelo vírus da Hepatite C, uma vez que isto não ocorre em infecções por outros vírus da mesma família. Esta persistência é relacionada às estratégias virais ativas e passivas de evasão e modulação do sistema imune (GALE et al., 2005), que envolvem a falha de células T vírus-específicas, o escape viral por mutações, supressão direta por citocinas ou células T regulatórias (KNOLLE et al., 2014). A progressão, a persistência e a alta prevalência da Hepatite C explicam a grande relevância desta doença para a saúde pública e os esforços na busca de tratamentos (APPEL et al., 2008; TREMBLING et al., 2013).

O vírus da Hepatite C (HCV) foi inicialmente isolado de chimpanzés infectados com sangue de pacientes que apresentaram a hepatite não-A e não-B após transfusão (CHOO et al., 1989). Atualmente, estima-se que aproximadamente 300 milhões de pessoas estejam infectadas pelo HCV e destes, 180 milhões de forma crônica (BROWN et al., 2003; CHAMBERS et al., 2005; EL-SERAG, 2012). Além disso, a cada ano aproximadamente 3 a 4 milhões de novos casos surgem e a distribuição global dos infectados não é homogênea (HUANG et al., 2006; MOHD HANAFIAH et al., 2013; SY et al., 2006) (**Figura 1**). No Brasil, a porcentagem de anticorpos anti-HCV em doadores de sangue detectados apresenta diferentes prevalências de acordo com a região: 0,8 a 5,9% no Norte, 0,7 a 1,7% no Nordeste, 1,4% no Centro-Oeste, 0,4 a 2,6% no Sudeste e 0,6 a 0,7% no Sul (MARTINS et al., 2011).

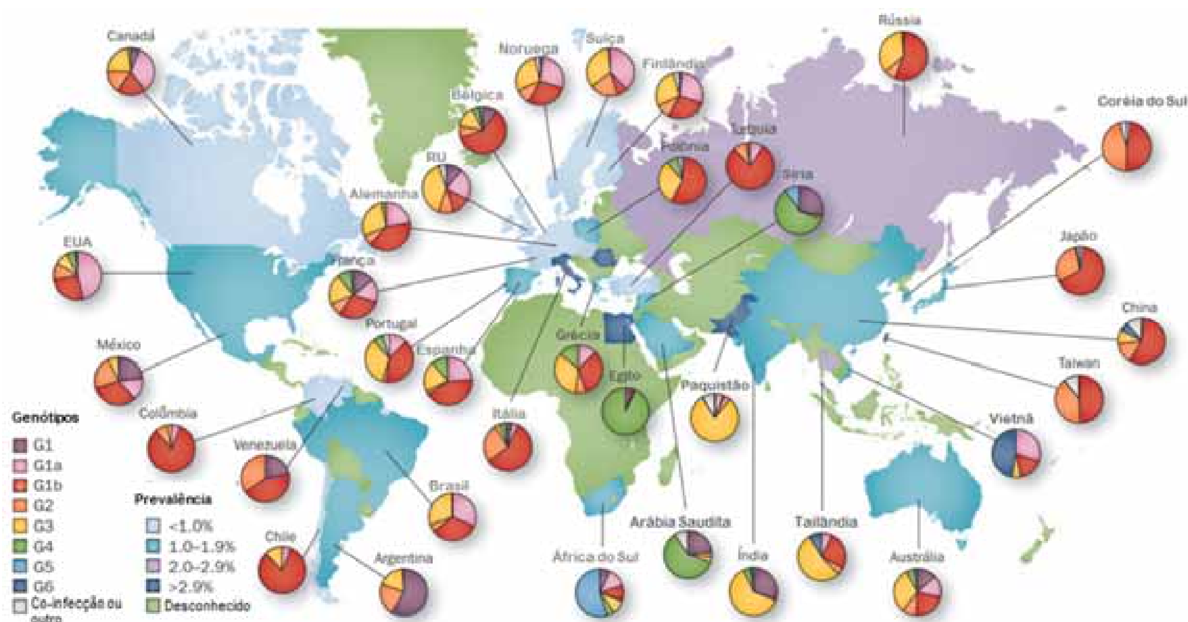


Figura 1. Prevalência do HCV e de cada um dos genótipos no mundo (HAJARIZADEH et al., 2013).

A transmissão do HCV é principalmente via parenteral. A partir de 1990, com a introdução de métodos para a detecção de anti-HCV em amostras de sangue, o número de novos casos de hepatite C por transfusão de sangue decresceu em vários países (SOLDAN et al., 2003). Atualmente, o uso de drogas intravenosas é a principal via identificada de transmissão (ESTEBAN et al., 2008; GREBELY et al., 2013).

Outras vias de transmissão também são estudadas, como a infecção ocupacional por meio de agulhas contaminadas, a qual resulta em soroconversão em aproximadamente 0 a 7% dos casos, representando um risco de transmissão intermediária entre o apresentado pelo HIV (0,3%) e pelo vírus da hepatite B (23 a 62%) (ESTEBAN et al., 2008; WASLEY et al., 2000).

A infecção do vírus por via sexual não é comprovada e estudos nesta área apresentam resultados contraditórios (DIENSTAG, 1997; FRANCO et al., 2014). As vias de transmissão vertical, ou seja, de mãe para filho acontecem principalmente em casos de mães imunocomprometidas, coinfectadas com HIV e usuárias de drogas intravenosas (PRASAD et al., 2013; SCHACKMAN et al., 2007). No entanto, estima-se que em mais de 40% dos casos

de infecção pelo HCV não são reconhecidos nenhum dos fatores de risco de transmissão (CLARKE et al., 2006; MORADPOUR et al., 2001).

1.2. O vírus da Hepatite C

O HCV é classificado na família *Flaviviridae*, gênero *Hepacivirus*, embora seja diferente em vários aspectos da organização do genoma em relação aos demais membros dessa família (SIMMONDS, 2004).

O RNA genômico deste vírus é uma fita simples de polaridade positiva, com aproximadamente 9.600 pares de bases. O genoma viral é composto por uma região não-traduzida 5\` (5\`UTR), uma região aberta de leitura que codifica uma poliproteína precursora de aproximadamente 3.000 aminoácidos e uma região não-traduzida 3\` (3´ UTR) (LIANG et al., 2007).

A 5' UTR é altamente conservada e possui um sítio de entrada de ribossomo (IRES), essencial para a tradução de RNA viral. Esta é seguida das regiões codificadoras das seguintes proteínas estruturais (**Figura 2**): "core" (Proteína C), Envelope 1 (E1) e Envelope 2 (E2); uma proteína associada à membrana (p7) e proteínas não-estruturais 2 (NS2), 3 (NS3), 4A (NS4A) e 4B (NS4B), 5A (NS5A) e 5B (NS5B). Seguido da NS5B, a 3' UTR é composta por uma pequena região variável, uma cauda poli (U/UC) e uma região altamente conservada. A poliproteína traduzida é posteriormente clivada por proteases virais e celulares, produzindo as proteínas estruturais e não-estruturais (LINDENBACH, 2011; LINDENBACH et al., 2005; LYRA et al., 2004; MORADPOUR et al., 2001).

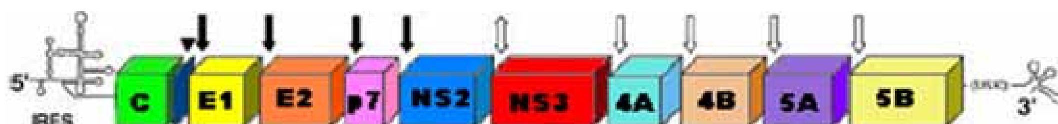


Figura 2. Esquema do genoma do HCV, demonstrando as proteínas codificadas "core" (Proteína C), Envelope 1 (E1) e Envelope 2 (E2), proteína associada à membrana (p7) e proteínas não-estruturais 2 (NS2), 3 (NS3), 4A (NS4A) e 4B (NS4B), 5A (NS5A) e 5B (NS5B) e regiões não codificadas como o sítio de entrada do ribossomo (IRES) e a cauda poli U/UC. As setas indicam as posições de clivagem (BARTENSCHLAGER et al., 2011).

1. 3. *Variabilidade do Genoma Viral e quasispecies*

O vírus da hepatite C é caracterizado pela grande diversidade na sequência de nucleotídeos entre isolados de pacientes diferentes e do mesmo paciente (CHOO et al., 1991; KATO et al., 1990). Essa diversidade é devido à ausência de atividade corretiva pela incorporação de bases errôneas da RNA polimerase RNA-dependente e rápida replicação (CARABALLO CORTES et al., 2013; MARTELL et al., 1992). A frequência média de mutações de nucleotídeos varia de 1,4 x 10^3 a 1,9 x 10^3 substituições por nucleotídeo por ano (LE GUILLOU-GUILLEMETTE et al., 2007). O genoma viral apresenta regiões altamente conservadas como as regiões não traduzidas 5’ e 3’, e de alta variabilidade como a E2 e a V3 da NS5A (Figura 3) (OKAMOTO et al., 1990; T. TANAKA et al., 1996; WEINER et al., 1991). Regiões com conservação intermediaria (core, E1 e NS5B) são utilizadas para determinar genótipos e subtipos (CHAYAMA et al., 1993; SIMMONDS et al., 1994). Atualmente, o HCV é classificado em 7 genótipos (denominados com números) que apresentam diferentes prevalências de acordo com a região geográfica. Os genótipos 1, 2 e 3 apresentam distribuição geográfica ampla (MCOMISH et al., 1994). E o genótipo 3 se destaca no sul da Ásia, Oceania, Europa, Canadá e alguns países da América do Sul, inclusive o Brasil (Figura 1)(CHAYAMA et al., 2011; HAJARIZADEH et al., 2013).

Os genótipos apresentam similaridade nucleotidica de 70 a 80%, os mais de 80 subtipos de 80 a 90% (KUIKEN et al., 2009; SIMMONDS, 2004). Nos indivíduos infectados, o HCV circula como uma população de variantes de mesma origem, que diferem em menos de 10% na sequência de nucleotídeos, denominadas *quasispescies* (DOMINGO et al., 2006).

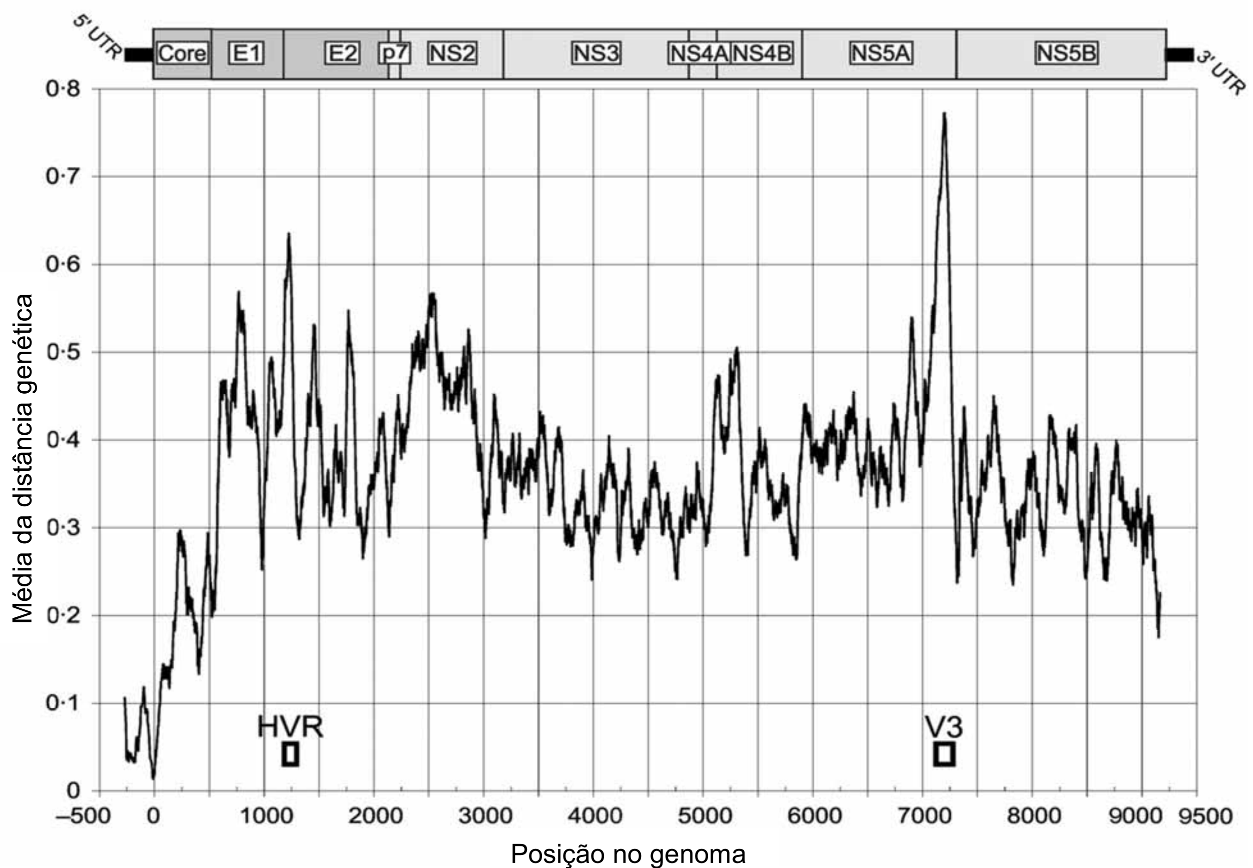


Figura 3. Gráfico de variabilidade genética ao longo do genoma de HCV. HVR – Região hipervariável, V3 – região variável 3 (FARCI, 2011).

As altas taxas de mutações em vírus de RNA, por um lado, podem ser deletérias para os vírus, já que resultam em partículas não infecciosas. No entanto, as altas taxas de mutações podem conferir às *quasispecies* virais maior probabilidade de se adaptarem ao novo ambiente e aos desafios da infecção (DOMINGO et al., 1997; EIGEN, 1993; EIGEN et al., 1988).

A adaptabilidade de replicação de um organismo em seu ambiente é definida pelo valor adaptativo ou *fitness*. Para os vírus de RNA, o *fitness* é estimado pela habilidade relativa de produzir uma progênie infecciosa estável em um dado ambiente. O acúmulo de mutações adquiridas pela replicação/seleção de *quasispecies* virais criam fenótipos diferentes que podem escapar das pressões seletivas que limitam sua replicação. Isso é um problema para o controle de doenças virais, pois reflete no escape de vacinas, nas mutações virais resistentes a drogas e na variação antigênica das populações de *quasispecies* (DOMINGO et al., 1997).

1.4. A Região Hipervariável 1 da Proteína do Envelope 2

Dentro do genoma do HCV, três regiões do gene codificante de E2 apresentam alta variabilidade (Figura 3), denominadas regiões hipervariáveis 1, 2 e 3 (respectivamente HVR1, HVR2 e HVR3) (LAPIERRE et al., 2011).

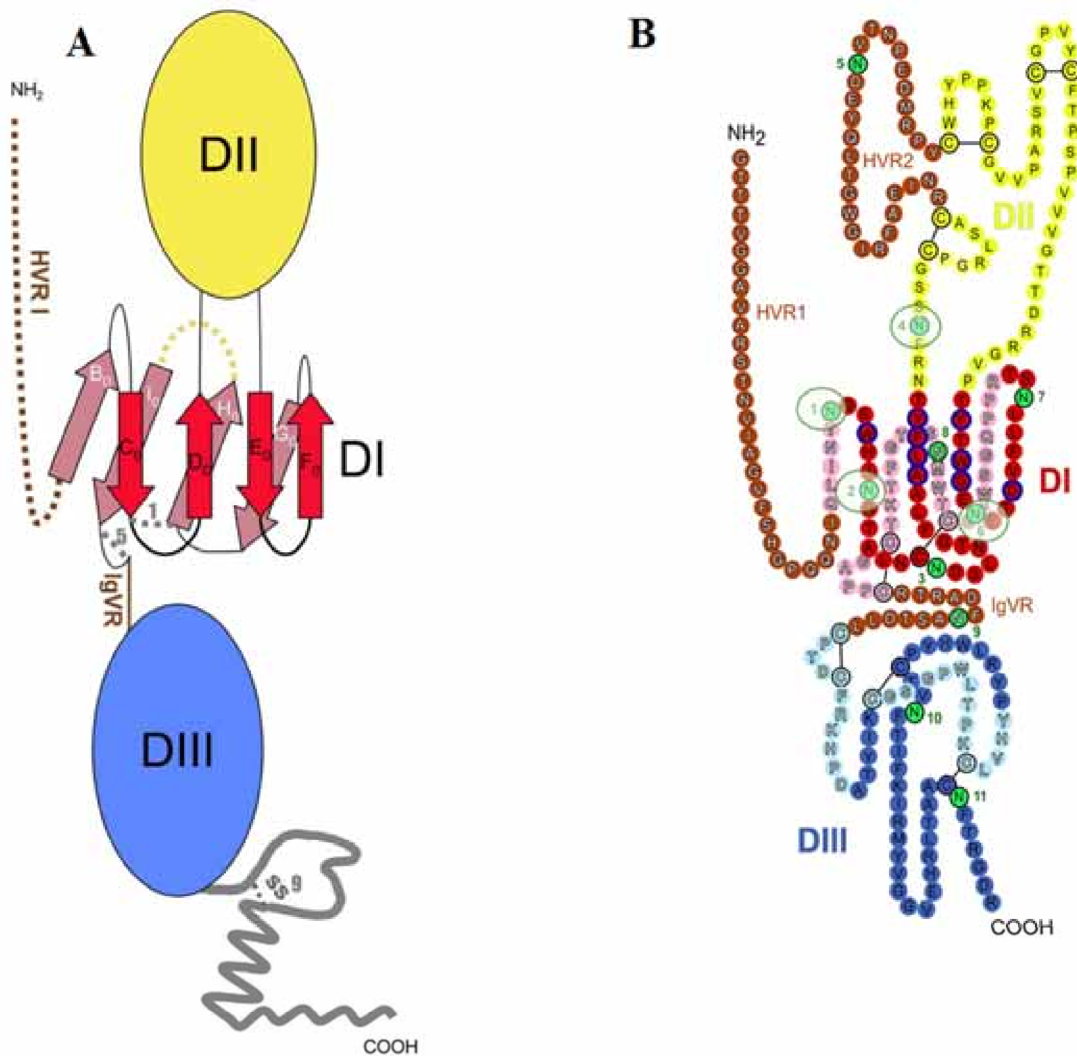


Figura 4. Esquema de E2, demonstrando os três domínios (A) e detalhe da sequência em B, salientando HVR1 e HVR2 (KREY et al., 2010).

A HVR1 consiste de uma sequência de 27 aminoácidos localizados no aminoterminal de E2 (resíduos 384-410 da poliproteína) (WEINER et al., 1991). A HVR2 localiza-se posteriormente, nos aminoácidos 474-482 (KATO et al., 1992a; KATO et al., 1992b). A

HVR3 posiciona-se entre HVR1 e HVR2 (resíduos 431-466) e foi o último a ser integrado ao modelo estrutural canônico da E2 (TROESCH et al., 2006). Os perfis de exposição ao solvente e resíduos específicos destas três regiões são condizentes com reconhecimento de células alvo, ligação da partícula viral e entrada do vírus na célula (LAW et al., 2008; PENIN et al., 2001; TORRES-PUENTE et al., 2008b). Considerando que as proteínas do envelope 1 e 2 (E1 e E2) do HCV são importantes alvos da resposta imune humoral e mediada por células do hospedeiro, as regiões hipervariáveis também sofrem grande pressão seletiva. A contínua variação de aminoácidos nestas regiões esta associado com fases de escape do sistema imune do hospedeiro, progressão e persistência da doença (LAPIERRE et al., 2011; SHOUKRY et al., 2004). Portanto, a HVR1 pode ser utilizada para identificar linhagens virais individuais e para o estudo de *quasispecies* em HCV (GAUDY et al., 2003). A predição de estrutura da HVR1 indica alta flexibilidade estrutural e grande área exposta (YAGNIK et al., 2000). Esta região ainda possui epítopos lineares de reconhecimentos por células B e sugere-se que seja o principal domínio imunogênico em E2 (WEINER et al., 1992). Além disso, vários estudos encontraram indícios de que a HVR1 apresenta papel funcional na ligação e entrada do HCV nas células-alvo: (i) anticorpos monoclonais anti-HVR1 inibem a ligação da partícula de HCV a células T humanas *in vitro* (Y. H. ZHOU et al., 2000); (ii) soro sanguíneo de coelhos hiperimunes ao carboxiterminal da HVR1 neutraliza a infectividade do HCV em chimpanzés (FARCI et al., 1996); e (iii) pequenas quantidades de RNA viral são detectadas em chimpanzés após infecção com HCV defectivo de HVR1 (FORNS et al., 2000).

A taxa de mutação na HVR1 é estimada em 0,1 a 0,2 nucleotídeos por sítio por genoma por ano, e várias mutações não-sinônimas ocorrem no período de um ano ou mais (ODEBERG et al., 1995). As *quasispecies* da HVR1 podem também estar associadas à resistência ao tratamento (GAUDY et al., 2003).

1.6. Tratamento e Tipos de Respostas ao Tratamento

A resposta ao tratamento antiviral do HCV pode ser classificada em três grupos: resposta virológica sustentada (SVR), resposta ao final do tratamento (RFT), e não-resposta (NR). A SVR é definida pela ausência de detecção do RNA de HCV ao término do tratamento e seis meses após este. Os pacientes com resposta transitória (RFT) apresentam ausência do RNA viral ao tratamento e detecção a partir de algum ponto após este. Os pacientes NR apresentam RNA de HCV detectado durante e após o término da terapia (FELD et al., 2005).

A monoterapia com Interferon (IFN) alfa forneceu o primeiro tratamento do vírus para os pacientes com hepatite C crônica. Essa substância não inibe uma função enzimática específica do vírus, mas modula as respostas do sistema imune e estabelece um estado antiviral inespecífico nas células infectadas pela ativação de diferentes genes celulares (MORENO et al., 2006), uma vez que os IFNs são citocinas chaves no estabelecimento de resposta antiviral, classificado em três tipos (tipo I, II e III) baseando-se no receptor destas moléculas, características estruturais e atividade biológica (LOPUSNA et al., 2013)

O IFN-α é classificado em IFN do tipo I, grupo que abrange citocinas essenciais nas respostas iniciais contra várias infecções virais. As células infectadas por vírus sintetizam e secretam os IFN do tipo I, que atua de forma autócrina e parácrina induzindo o estado antiviral nas células-hospedeiro (PENG et al., 2008). Desta forma, o IFN é expresso em uma variedade de tipos celulares, levando a expressão de um conjunto de genes envolvidos na apoptose, anti-crescimento e ativação do sistema imune inato e adaptativo (PERRY et al., 2005; UZE et al., 2007)

Para melhorar a farmacocinética do Interferon, formas peguiladas foram desenvolvidas, aumentando a meia-vida da droga no organismo. Até 2010, a terapia recomendada pela Organização Mundial de Saúde (OMS) tratava-se de uma combinação de

Peginterferon e Ribavirina (FRIED et al., 2004). Essa terapia combinada é efetiva em apenas uma fração dos pacientes. A SVR é alcançada em aproximadamente 50% dos indivíduos infectados com HCV genótipo 1 e em 70% para os genótipos 2 e 3 (CONTEDUCA et al., 2013; HUANG et al., 2006; KLENERMAN et al., 2009).

A Ribavirina é um análogo de nucleosídeo com atividade antiviral de amplo espectro, descoberta em 1970 (SIDWELL et al., 1972; WITKOWSKI et al., 1972). Para o tratamento da Hepatite C, este composto é utilizado desde 1998 em combinação com o IFN. Embora o mecanismo de inibição da Ribavirina para o HCV não seja totalmente elucidado, quatro principais mecanismos são propostos. Duas ações indiretas: (1) estimulação da imunidade antiviral do hospedeiro mediada por células T, pela conversão doa resposta Th do tipo 2 para o tipo 1 (TAM et al., 1999) e (2) inibição da enzima inosina monofosfato desidrogenase (IMPDH), uma enzima celular necessária na síntese de guanosina trifostato (S. ZHOU et al., 2003). Em adição, a Ribavirina pode apresentar ação direta anti-HCV: (1) inibição da NS5B polimerase, atuando como competidor de substrato ou terminação pré-matura da fita em síntese (MAAG et al., 2001) e (2) atuando como mutagênico, levando a rápida mutação do RNA viral além do limite de “erro de catástrofe” e aumentando a taxa de mutações letais (CROTTY et al., 2001; LAU et al., 2002; LIU et al., 2012).

As primeiras drogas antivirais de ação direcionada (DAAs) inibidoras da atividade da NS3/4A para Hepatite C crônica foram implementadas no final de 2010 nos Estados Unidos. No Brasil, os inibidores para HCV foram introduzidos no final de 2011 (VELOSA et al., 2012)Atualmente, para infecções do genótipo 1, a indicação da Organização Mundial de Saúde (OMS) é a utilização de terapia combinada tripla: IFN+Ribavirina+Telaprevir ou Boceprevir. Esta nova terapia, chega a alcançar até 70-90% de SVR em pacientes sem tratamento anterior com PegIFN (LAWITZ et al., 2013). Com a introdução de tratamentos HCV genótipo 1 específicos, a atenção de clínicos e pesquisadores voltaram-se aos 50-70

milhões de pacientes infectados com outros genótipos. Neste contexto, genótipo 3 agora apresenta a menor taxa de SVR ao tratamento (65-70%) e consequentemente o genótipo mais complicado de se tratar (TAPPER et al., 2013; WARTELLE-BLADOU et al., 2012).

1.7. Fatores preditivos de resposta ao tratamento baseado em IFN

A baixa susceptibilidade (resposta) a tratamentos antivirais e a alta taxa de persistência das infecções provavelmente estão relacionados a uma complexa relação entre fatores virais e do hospedeiro (TORRES-PUENTE et al., 2008a). Os fatores virais incluem o genótipo do HCV, nível de replicação (carga viral) e complexidade genética da população de *quasispecies* antes do início do tratamento (MORENO et al., 2006).

Como discutido anteriormente, o tratamento baseado em administração de IFN apresenta sucesso em aproximadamente 50-70% dos pacientes, dependendo do genótipo. Além disso, o tratamento com PegIFN+RBV é financeiramente custoso e apresenta diversos efeitos colaterais. Desta forma, considera-se de grande importância os estudos de fatores de predição de resposta ao tratamento. Estes fatores podem contribuir para a escolha de qual o tipo de tratamento mais indicado para alcançar a SVR em diferentes casos (BACKUS et al., 2007; KIM et al., 2012).

Várias metodologias têm sido aplicadas para esta finalidade. Um dos mais confiáveis é o monitoramento da carga viral na fase inicial do tratamento (DI BISCEGLIE et al., 2007). Entretanto, até o momento não ha nenhuma metodologia estabelecida que seja totalmente confiável para determinar a resposta antes do início do tratamento (KUROSAKI et al., 2011). A seguir, detalham-se alguns fatores do hospedeiro e do vírus estudados com relação à resposta ao IFN.

Embora a associação de características do paciente como idade, sexo, etnia, nível de plaquetas, fibrose hepática, obesidade, co-infecção com HBV/HIV e alcoolismo já foram

determinados há alguns anos (KUROSAKI et al., 2011), apenas recentemente foram identificados fatores genéticos associados à resposta a terapia. Vários alelos na região codificante da interleucina 28B (IL28B), no cromossomo 19 (DOEHRING et al., 2010; RAUCH et al., 2010; Y. TANAKA et al., 2009), foram relacionados à resposta ao tratamento e regulação de genes ligados a expressão de IFN intra-hepático (BACKUS et al., 2007; DAVIS et al., 2003; RAU et al., 2012). Estudos de diferentes localidades (Europa, Ásia, Austrália, África), mostraram essa correlação, indicando que estes polimorfismos são aplicáveis a diferentes etnias (ALESTIG et al., 2011; CAVALCANTE et al., 2012; SHI et al., 2012; THOMAS et al., 2009).

Apesar da associação entre tratamento e IL28B serem melhores descritos e já praticados na rotina clinica, outros polimorfismos foram apontados como relacionados a essa resposta, como genes relacionados a transportadores de ribavirina, inosina trifosfatase (ITPA), receptor de vitamina D e bomba exportadora de sal biliar (Bsep). Desta forma, os marcadores genéticos do hospedeiro são importantes na decisão da terapia mais adequada e na prevenção da progressão da hepatite C (RAU et al., 2012).

Assim como características do paciente, vários fatores virais associados à resposta ao tratamento foram identificados, embora não totalmente elucidados. A variabilidade viral é um destes fatores e confere potencial adaptativo ao vírus, o qual sugere estar implicado na evasão e no controle da resposta do hospedeiro à infecção. Como discutido anteriormente, a sensibilidade e o comportamento à terapia com IFN são diferentes entre os genótipos (QASHQARI et al., 2013). Além disso, o ambiente hostil do hospedeiro pode levar a seleção de variantes resistentes de HCV a partir de um *pool* de *quasispecies* pré-existentes e/ou por adaptação genética viral. De fato, estudos com sequenciamento mostraram que a remissão da infecção aguda está associada a grande redução da complexidade das *quasispecies* virais em regiões específicas do genoma viral, enquanto a progressão para a infecção crônica e a

resistência à terapia com IFN está associada ao aumento da complexidade genética viral (GALE et al., 2005; JARDIM et al., 2009).

Diversas proteínas do HCV estão relacionadas a não resposta ao tratamento com IFN, especiamente demonstradas no genótipo 1. Dentre estas as mais detalhadas são as proteínas do capsídeo, E2, NS3 e NS5A (QASHQARI et al., 2013).

A primeira observação desta relação em E2 ocorreu em 1999, em que a presença desta proteína bloqueou a fosforilação da PKR, conferindo escape ao efeito antiviral do IFN (CHAYAMA et al., 2000; TAYLOR et al., 1999). Também foi observado que a fosforilação em resíduos de serina de E2 pode estar relacionada a resistência ao IFN (AFZAL et al., 2011).

Como citado anteriormente, a proteína E2 possui epítopos de reconhecimento do sistema imune como a HVR1 e características desta região podem estar relacionadas a resposta ao tratamento (GAUDY et al., 2003). Os polimorfismos conformacionais (POLYAK et al., 1998) e heterogeneidade da HVR1 (OKADA et al., 1992; TOYODA et al., 1997) podem ser preditivos da resposta ao IFN. Sugerindo que variantes presentes antes do início do tratamento estão intrinsicamente associadas a resistência ao tratamento (GAUDY et al., 2003). Esta relação pode resultar em modelos preditivos de resposta ao tratamento a partir de análises da HVR1 (LARA et al., 2011).

1.8. Sequenciamento de nova geração aplicado a Virologia

O sequenciamento dideoxinucleotídico (Sanger) de produtos não-clonados do cDNA de vírus do plasma sanguíneo é uma técnica bastante utilizada na detecção de mutações de resistência a drogas antivirais (WANG et al., 2007). Entretanto, a principal limitação do sequenciamento direto é a incapacidade de detectar mutações em baixa frequência, numa população viral heterogênea que coexiste na amostra de plasma do paciente (PALMER et al.,

2005). Diversos estudos mostraram que variantes resistentes à drogas presentes em número inferior na amostra, não são detectadas pelo sequenciamento direto e são geralmente responsáveis pelo insucesso de um novo regime de tratamento antiviral (JOURDAIN et al., 2004; KAPOOR et al., 2004; LECOSSIER et al., 2005; PALMER et al., 2005). Para contornar esta limitação, diversas abordagens vêm sendo desenvolvidas para detectar estes mutantes, entretanto, nenhuma destas comprovou sua eficiência em estudos clínicos (WANG et al., 2007).

A plataforma de sequenciamento *454 Life Sciences GS* permite a amplificação paralela em grande escala, utilizando volumes de picolitros e o pirossequenciamento de moléculas de DNA individuais de tamanho até 300 pares de bases (MARGULIES et al., 2005). Simons e colaboradores descreveram dois casos em que o pirossequenciamento de alta capacidade (UDPS – do Inglês "Ultra-Deep PyroSequencing") detectou variantes em minoria com mutações de resistência a antivirais, em pacientes anteriormente tratados. Tais mutações já não eram mais detectáveis pelo sequenciamento direto do PCR (ROZERA et al., 2012). O estudo de Wang e colaboradores, ao analisar sequências de duas proteínas do HIV-1, demonstrou que o UDPS detectou uma média de 58 variantes por amostra, enquanto que sequenciamento direto apresentou média de 8 variantes/amostra (WANG et al., 2007).

Além disso, o UDPS é uma técnica mais simples e com custo mais baixo do que o sequenciamento a partir de clonagem convencional, uma vez que a clonagem pelo UDPS é realizada automaticamente por um processo de emulsão em água-em-óleo, processo que não necessita de transformação em bactéria ou de inúmeras diluições e reações de PCR. O custo por base é 10% menor do que o custo gerado pelo sequenciamento Sanger, tornando possível o sequenciamento de cada nucleotídeo de 100 vezes ou até 1000 vezes (SHENDURE et al., 2004). O sequenciador do modelo GS20 tem sido substituído por outra máquina de segunda

geração (FLX Sequencer), que consegue ler fragmentos de até 400nucleotídeos e produz o dobro de leituras gerando os mesmos custos do sequenciador GS (LEAMON et al., 2007).

Atualmente, o UDPS é utilizado para a identificação de variantes resistentes às drogas, além de diversas outras análises. Dentre os estudos do primeiro tipo, podemos citar os trabalhos de Simen e colaboradores (2009) e Margeridon-Thermet e colaboradores (2009), que por meio do UDPS detectaram um número diferenciado de pacientes que possuíam variantes do HIV e HBV, respectivamente, com resistência a drogas. Esta técnica é, portanto, um importante aliado na predição de resposta ao tratamento (MARGERIDON-THERMET et al., 2009; SIMEN et al., 2009). Embora seja ainda uma área de estudo recente, o UDPS já é utilizado em investigações sobre a diversidade genética e epidemiológica do HCV (CARABALLO CORTES et al., 2013). Além disso, o sequenciamento de segunda geração é aplicado em estudos sobre a transmissão do vírus (ESCOBAR-GUTIERREZ et al., 2012), dinâmica viral ao longo da infecção (BULL et al., 2011) e detecção de mutantes resistentes a antivirais (FONSECA-CORONADO et al., 2012).

O UDPS, desta forma, apresenta versatilidade, alta eficiência e baixo custo por sequência, o que explica o crescente uso desta ferramenta em investigações de diversas áreas (CUMMINGS et al., 2013).

2. Justificativa

A Hepatite C é uma doença que incide em todos os países, inclusive no Brasil, onde se estima que 0,8 – 3,4% da população esteja infectada com o agente viral, denominado vírus da Hepatite C (HCV). Os indivíduos infectados podem não apresentar os sintomas da doença e em aproximadamente 80% dos casos, a doença progride em hepatite crônica.

Atualmente, não há vacinas aprovadas por testes clínicos. O tratamento mais recente contra o HCV baseia-se na administração de antivirais específicos para o HCV, entretanto disponíveis apenas para o genótipo 1 no Brasil. Os demais genótipos, incluindo o genótipo 3, são tratados de forma tradicional com IFN e Ribavirina, eficaz em aproximadamente 80% dos pacientes infectados. No cenário recente, o genótipo 3 apresenta a maior probabilidade de resistência ao tratamento e possui ampla distribuição geográfica. A resistência viral e a resposta ao tratamento podem estar relacionadas a fatores do hospedeiro e do vírus, como a variabilidade do genoma viral.

Uma hipótese para o alto número de casos resistentes é que a terapia pode gerar uma pressão de seleção nas variantes presentes no paciente, permanecendo no indivíduo apenas as quasispecies que apresentem resistência ao medicamento e ao sistema imune do hospedeiro, mesmo que esta variante seja minoria inicialmente. Desta forma, a composição de quasispecies do hospedeiro possui um papel crucial na determinação da resposta ao tratamento.

A região hipervariável 1 (HVR1) participa da entrada do vírus na célula e na interação com anticorpos. Neste caso, a alta variabilidade desta região determina o escape do sistema imune e persistência da doença. Além disso, vários estudos apontam para uma relação entre heterogeneidade genética da HVR1 e resposta ao tratamento.

O pirossequenciamento de alta capacidade (UDPS) apresenta diversas vantagens sobre ambas as técnicas tradicionais de sequenciamento. Além disso, demonstra alta sensibilidade na detecção de mutações pontuais para o estudo de resistência à antivirais de diversos vírus de doenças crônicas, como o HIV, o HBV e o HCV. O UDPS gera dados detalhados sobre as variantes presentes em uma determinada amostra que podem ser relacionadas à resistência viral ao tratamento.

Por fim, estudos sobre resistência viral ao tratamento são essenciais para que no futuro seja possível a predição de resposta uma terapia, bem como a escolha do tratamento mais adequado para cada paciente. Desta forma, será possível oferecer melhor qualidade de tratamento aos infectados com HCV e diminuir os custos do governo com o tratamento de não-respondedores e respondedores ao final do tratamento.

3. Objetivos

3.1. Objetivo Geral

Analisar a composição detalhada de *quasispecies* da HVR1 do HCV em amostras pré-tratamento de pacientes que já foram tratados e apresentaram diferentes tipos de resposta ao tratamento com Interferon e Ribavirina.

3.2. Objetivos específicos

1. Obter as sequências da região codificante da HVR1 do HCV de amostras do genótipo 3 de pacientes que mostraram diferentes respostas ao tratamento segundo a técnica de UDPS;
2. Analisar a possível associação entre sítios de mutação e resposta ao tratamento;
3. Identificar polimorfismos de base única (SNP) na HVR1 associados à resposta ao tratamento;
4. Analisar a estrutura populacional por estatística Bayesiana;
5. Identificar alterações na sequência de aminoácidos, físico-químicas e estruturais na HVR1;
6. Predizer possíveis sítios de reconhecimentos por células B e as alterações nas variantes da HVR1 encontradas em cada paciente.

4. Referências Bibliográficas

AFZAL, S. et al. Envelope 2 protein phosphorylation sites S75 & 277 of hepatitis C virus genotype 1a and interferon resistance: a sequence alignment approach. **Virol J**, v. 8, n. p. 71, 2011.

ALESTIG, E. et al. Core mutations, IL28B polymorphisms and response to peginterferon/ribavirin treatment in Swedish patients with hepatitis C virus genotype 1 infection. **BMC Infect Dis**, v. 11, n. 1, p. 124, 2011.

ALTER, H.J. et al. Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. **N Engl J Med**, v. 321, n. 22, p. 1494-1500, 1989.

APPEL, N. et al. Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. **PLoS Pathog**, v. 4, n. 3, p. e1000035, 2008.

BACKUS, L.I. et al. Predictors of response of US veterans to treatment for the hepatitis C virus. **Hepatology**, v. 46, n. 1, p. 37-47, 2007.

BARTENSCHLAGER, R. et al. Assembly of infectious hepatitis C virus particles. **Trends Microbiol**, v. 19, n. 2, p. 95-103, 2011.

BROWN, R.S., JR. & GAGLIO, P.J. Scope of worldwide hepatitis C problem. **Liver transplantation : official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society**, v. 9, n. 11, p. S10-13, 2003.

BULL, R.A. et al. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. **PLoS Pathog**, v. 7, n. 9, p. e1002243, 2011.

CARABALLO CORTES, K. et al. Ultradeep pyrosequencing of hepatitis C virus hypervariable region 1 in quasispecies analysis. **Biomed Res Int**, v. 2013, n. p. 626083, 2013.

CAVALCANTE, L.N. et al. IL28B polymorphisms are markers of therapy response and are influenced by genetic ancestry in chronic hepatitis C patients from an admixed population. **Liver Int**, v. 32, n. 3, p. 476-486, 2012.

CHAMBERS, T.J. et al. Quasispecies heterogeneity within the E1/E2 region as a pretreatment variable during pegylated interferon therapy of chronic hepatitis C virus infection. **Journal of virology**, v. 79, n. 5, p. 3071-3083, 2005.

CHAYAMA, K. et al. Genotypic subtyping of hepatitis C virus. **J Gastroenterol Hepatol**, v. 8, n. 2, p. 150-156, 1993.

CHAYAMA, K. et al. Association of amino acid sequence in the PKR-eIF2 phosphorylation homology domain and response to interferon therapy. **Hepatology**, v. 32, n. 5, p. 1138-1144, 2000.

CHAYAMA, K. & HAYES, C.N. Hepatitis C virus: How genetic variability affects pathobiology of disease. **J Gastroenterol Hepatol**, v. 26 Suppl 1, n. p. 83-95, 2011.

CHOO, Q.L. et al. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. **Science**, v. 244, n. 4902, p. 359-362, 1989.

CHOO, Q.L. et al. Genetic organization and diversity of the hepatitis C virus. **Proc Natl Acad Sci U S A**, v. 88, n. 6, p. 2451-2455, 1991.

CLARKE, A. & KULASEGARAM, R. Hepatitis C transmission -- where are we now? **Int J STD AIDS**, v. 17, n. 2, p. 74-80; quiz 80, 2006.

CONTEDUCA, V. et al. Therapy of chronic hepatitis C virus infection in the era of direct-acting and host-targeting antiviral agents. **J Infect**, v. n. p. 2013.

CROTTY, S. et al. RNA virus error catastrophe: direct molecular test by using ribavirin. **Proc Natl Acad Sci U S A**, v. 98, n. 12, p. 6895-6900, 2001.

CUMMINGS, P.J. et al. Pyrosequencing for microbial identification and characterization. **J Vis Exp**, v. n. 78, p. 2013.

DAVIS, G.L. et al. Early virologic response to treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C. **Hepatology**, v. 38, n. 3, p. 645-652, 2003.

DI BISCEGLIE, A.M. et al. Early virologic response after peginterferon alpha-2a plus ribavirin or peginterferon alpha-2b plus ribavirin treatment in patients with chronic hepatitis C. **J Viral Hepat**, v. 14, n. 10, p. 721-729, 2007.

DIENSTAG, J.L. Sexual and perinatal transmission of hepatitis C. **Hepatology**, v. 26, n. 3 Suppl 1, p. 66S-70S, 1997.

DOEHRING, A. et al. Screening for IL28B gene variants identifies predictors of hepatitis C therapy success. **Antivir Ther**, v. 15, n. 8, p. 1099-1106, 2010.

DOMINGO, E. & HOLLAND, J.J. RNA virus mutations and fitness for survival. **Annual review of microbiology**, v. 51, n. p. 151-178, 1997.

DOMINGO, E. et al. Viruses as quasispecies: biological implications. **Curr Top Microbiol Immunol**, v. 299, n. p. 51-82, 2006.

EIGEN, M. et al. Statistical geometry in sequence space: a method of quantitative comparative sequence analysis. **Proc Natl Acad Sci U S A**, v. 85, n. 16, p. 5913-5917, 1988.

EIGEN, M. The origin of genetic information: viruses as models. **Gene**, v. 135, n. 1-2, p. 37-47, 1993.

EL-SERAG, H.B. Epidemiology of viral hepatitis and hepatocellular carcinoma. **Gastroenterology**, v. 142, n. 6, p. 1264-1273 e1261, 2012.

ESCOBAR-GUTIERREZ, A. et al. Identification of hepatitis C virus transmission using a next-generation sequencing approach. **J Clin Microbiol**, v. 50, n. 4, p. 1461-1463, 2012.

ESTEBAN, J.I. et al. The changing epidemiology of hepatitis C virus infection in Europe. **Journal of hepatology**, v. 48, n. 1, p. 148-162, 2008.

FARCI, P. et al. Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein. **Proc Natl Acad Sci U S A**, v. 93, n. 26, p. 15394-15399, 1996.

FARCI, P. New insights into the HCV quasispecies and compartmentalization. **Semin Liver Dis**, v. 31, n. 4, p. 356-374, 2011.

FELD, J.J. & HOOFNAGLE, J.H. Mechanism of action of interferon and ribavirin in treatment of hepatitis C. **Nature**, v. 436, n. 7053, p. 967-972, 2005.

FONSECA-CORONADO, S. et al. Specific detection of naturally occurring hepatitis C virus mutants with resistance to telaprevir and boceprevir (protease inhibitors) among treatment-naive infected individuals. **J Clin Microbiol**, v. 50, n. 2, p. 281-287, 2012.

FORNS, X. et al. Hepatitis C virus lacking the hypervariable region 1 of the second envelope protein is infectious and causes acute resolving or persistent infection in chimpanzees. **Proc Natl Acad Sci U S A**, v. 97, n. 24, p. 13318-13323, 2000.

FRANCO, S. et al. Detection of a Sexually Transmitted HCV Protease Inhibitor-Resistance Variant in a HIV-infected Homosexual Man. **Gastroenterology**, v. n. p. 2014.

FRIED, M.W. & HADZIYANNIS, S.J. Treatment of chronic hepatitis C infection with peginterferons plus ribavirin. **Seminars in liver disease**, v. 24 Suppl 2, n. p. 47-54, 2004.

GALE, M., JR. & FOY, E.M. Evasion of intracellular host defence by hepatitis C virus. **Nature**, v. 436, n. 7053, p. 939-945, 2005.

GAUDY, C. et al. Significance of pretreatment analysis of hepatitis C virus genotype 1b hypervariable region 1 sequences to predict antiviral outcome. **J Clin Microbiol**, v. 41, n. 8, p. 3615-3622, 2003.

GREBELY, J. et al. Moving the agenda forward: the prevention and management of hepatitis C virus infection among people who inject drugs. **Clin Infect Dis**, v. 57 Suppl 2, n. p. S29-31, 2013.

HAJARIZADEH, B. et al. Epidemiology and natural history of HCV infection. **Nat Rev Gastroenterol Hepatol**, v. 10, n. 9, p. 553-562, 2013.

HUANG, R.H. & HU, K.Q. A practical approach to managing patients with HCV infection. **International journal of medical sciences**, v. 3, n. 2, p. 63-68, 2006.

JARDIM, A.C. et al. Quasispecies of hepatitis C virus genotype 1 and treatment outcome with peginterferon and ribavirin. **Infect Genet Evol**, v. 9, n. 4, p. 689-698, 2009.
JOURDAIN, G. et al. Intrapartum exposure to nevirapine and subsequent maternal responses to nevirapine-based antiretroviral therapy. **N Engl J Med**, v. 351, n. 3, p. 229-240, 2004.
KAPOOR, A. et al. Sequencing-based detection of low-frequency human immunodeficiency virus type 1 drug-resistant mutants by an RNA/DNA heteroduplex generator-tracking assay. **J Virol**, v. 78, n. 13, p. 7112-7123, 2004.
KATO, N. et al. Sequence diversity of hepatitis C viral genomes. **Mol Biol Med**, v. 7, n. 6, p. 495-501, 1990.
KATO, N. et al. Characterization of hypervariable regions in the putative envelope protein of hepatitis C virus. **Biochem Biophys Res Commun**, v. 189, n. 1, p. 119-127, 1992a.
KATO, N. et al. Marked sequence diversity in the putative envelope proteins of hepatitis C viruses. **Virus Res**, v. 22, n. 2, p. 107-123, 1992b.
KIM, S.R. et al. Prediction of response to pegylated interferon/ribavirin combination therapy for chronic hepatitis C genotype 1b and high viral load. **J Gastroenterol**, v. 47, n. 10, p. 1143-1151, 2012.
KLENERMAN, P. et al. What are the prospects for controlling hepatitis C? **PLoS Med**, v. 6, n. 6, p. e1000096, 2009.
KNOLLE, P.A. & THIMME, R. Hepatic immune regulation and its involvement in viral hepatitis infection. **Gastroenterology**, v. 146, n. 5, p. 1193-1207, 2014.
KREY, T. et al. The disulfide bonds in glycoprotein E2 of hepatitis C virus reveal the tertiary organization of the molecule. **PLoS Pathog**, v. 6, n. 2, p. e1000762, 2010.
KUIKEN, C. & SIMMONDS, P. Nomenclature and numbering of the hepatitis C virus. **Methods Mol Biol**, v. 510, n. p. 33-53, 2009.
KUROSAKI, M. et al. Pretreatment prediction of response to peginterferon plus ribavirin therapy in genotype 1 chronic hepatitis C using data mining analysis. **J Gastroenterol**, v. 46, n. 3, p. 401-409, 2011.
LAPIERRE, P. et al. Structural basis for broad neutralization of hepatitis C virus quasispecies. **PLoS One**, v. 6, n. 10, p. e26981, 2011.
LARA, J. et al. Coordinated evolution among hepatitis C virus genomic sites is coupled to host factors and resistance to interferon. **In Silico Biol**, v. 11, n. 5-6, p. 213-224, 2011.
LAU, J.Y. et al. Mechanism of action of ribavirin in the combination treatment of chronic HCV infection. **Hepatology**, v. 35, n. 5, p. 1002-1009, 2002.
LAW, M. et al. Broadly neutralizing antibodies protect against hepatitis C virus quasispecies challenge. **Nat Med**, v. 14, n. 1, p. 25-27, 2008.
LAWITZ, E. et al. Sofosbuvir for previously untreated chronic hepatitis C infection. **N Engl J Med**, v. 368, n. 20, p. 1878-1887, 2013.
LE GUILLOU-GUILLEMETTE, H. et al. Genetic diversity of the hepatitis C virus: impact and issues in the antiviral therapy. **World journal of gastroenterology : WJG**, v. 13, n. 17, p. 2416-2426, 2007.
LEAMON, J.H. & ROTHBERG, J.M. Cramming more sequencing reactions onto microreactor chips. **Chem Rev**, v. 107, n. 8, p. 3367-3376, 2007.
LECOSSIER, D. et al. Detection of minority populations of HIV-1 expressing the K103N resistance mutation in patients failing nevirapine. **J Acquir Immune Defic Syndr**, v. 38, n. 1, p. 37-42, 2005.
LIANG, Y. et al. Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded. **Biochemistry**, v. 46, n. 41, p. 11550-11558, 2007.
LINDENBACH, B.D. et al. Complete replication of hepatitis C virus in cell culture. **Science**, v. 309, n. 5734, p. 623-626, 2005.
LINDENBACH, B.D. Understanding how hepatitis C virus builds its unctuous home. **Cell Host Microbe**, v. 9, n. 1, p. 1-2, 2011.

LIU, W.L. et al. Ribavirin enhances the action of interferon-alpha against hepatitis C virus by promoting the p53 activity through the ERK1/2 pathway. **PLoS One**, v. 7, n. 9, p. e43824, 2012.

LOPUSNA, K. et al. Interferons lambda, new cytokines with antiviral activity. **Acta Virol**, v. 57, n. 2, p. 171-179, 2013.

LYRA, A.C. et al. Molecular biology and clinical implication of hepatitis C virus. **Braz J Med Biol Res**, v. 37, n. 5, p. 691-695, 2004.

MAAG, D. et al. Hepatitis C virus RNA-dependent RNA polymerase (NS5B) as a mediator of the antiviral activity of ribavirin. **J Biol Chem**, v. 276, n. 49, p. 46094-46098, 2001.

MARGERIDON-THERMET, S. et al. Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naive patients. **J Infect Dis**, v. 199, n. 9, p. 1275-1285, 2009.

MARGULIES, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. **Nature**, v. 437, n. 7057, p. 376-380, 2005.

MARTELL, M. et al. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. **J Virol**, v. 66, n. 5, p. 3225-3229, 1992.

MARTINS, T. et al. Epidemiologia da infecção pelo vírus da hepatite C. **Rev Assoc Med Bras**, v. 57, n. 01, p. 5, 2011.

MCOMISH, F. et al. Geographical distribution of hepatitis C virus genotypes in blood donors: an international collaborative survey. **J Clin Microbiol**, v. 32, n. 4, p. 884-892, 1994.

MOHD HANAFIAH, K. et al. Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. **Hepatology**, v. 57, n. 4, p. 1333-1342, 2013.

MORADPOUR, D. et al. Hepatitis C: an update. **Swiss Med Wkly**, v. 131, n. 21-22, p. 291-298, 2001.

MORENO, M.P. et al. Evidence of recombination in quasispecies populations of a Hepatitis C Virus patient undergoing anti-viral therapy. **Virology journal**, v. 3, n. p. 87, 2006.

ODEBERG, J. et al. Dynamic analysis of heterogeneous hepatitis C virus populations by direct solid-phase sequencing. **J Clin Microbiol**, v. 33, n. 7, p. 1870-1874, 1995.

OKADA, S. et al. The degree of variability in the amino terminal region of the E2/NS1 protein of hepatitis C virus correlates with responsiveness to interferon therapy in viremic patients. **Hepatology**, v. 16, n. 3, p. 619-624, 1992.

OKAMOTO, H. et al. The 5'-terminal sequence of the hepatitis C virus genome. **Jpn J Exp Med**, v. 60, n. 3, p. 167-177, 1990.

PALMER, S. et al. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. **J Clin Microbiol**, v. 43, n. 1, p. 406-413, 2005.

PENG, T. et al. Independent and cooperative antiviral actions of beta interferon and gamma interferon against herpes simplex virus replication in primary human fibroblasts. **J Virol**, v. 82, n. 4, p. 1934-1945, 2008.

PENIN, F. et al. Conservation of the conformation and positive charges of hepatitis C virus E2 envelope glycoprotein hypervariable region 1 points to a role in cell attachment. **J Virol**, v. 75, n. 12, p. 5703-5710, 2001.

PERRY, A.K. et al. The host type I interferon response to viral and bacterial infections. **Cell Res**, v. 15, n. 6, p. 407-422, 2005.

POLYAK, S.J. et al. Evolution of hepatitis C virus quasispecies in hypervariable region 1 and the putative interferon sensitivity-determining region during interferon therapy and natural infection. **J Virol**, v. 72, n. 5, p. 4288-4296, 1998.

PRASAD, M.R. & HONEGGER, J.R. Hepatitis C virus in pregnancy. **Am J Perinatol**, v. 30, n. 2, p. 149-159, 2013.

QASHQARI, H. et al. Understanding the molecular mechanism(s) of hepatitis C virus (HCV) induced interferon resistance. **Infect Genet Evol**, v. 19C, n. p. 113-119, 2013.

RAU, M. et al. Host genetic variants in the pathogenesis of hepatitis C. **Viruses**, v. 4, n. 12, p. 3281-3302, 2012.

RAUCH, A. et al. Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. **Gastroenterology**, v. 138, n. 4, p. 1338-1345, 1345 e1331-1337, 2010.

ROZERA, G. et al. Ultra-deep sequencing reveals hidden HIV-1 minority lineages and shifts of viral population between the main cellular reservoirs of the infection after therapy interruption. **J Med Virol**, v. 84, n. 6, p. 839-844, 2012.

SAITO, I. et al. Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. **Proc Natl Acad Sci U S A**, v. 87, n. 17, p. 6547-6549, 1990.

SCHACKMAN, B.R. & ONEDA, K. Preventing vertical hepatitis C virus transmission. **Clin Infect Dis**, v. 45, n. 6, p. 802, 2007.

SHENDURE, J. et al. Advanced sequencing technologies: methods and goals. **Nat Rev Genet**, v. 5, n. 5, p. 335-344, 2004.

SHI, X. et al. IL28B genetic variation is associated with spontaneous clearance of hepatitis C virus, treatment response, serum IL-28B levels in Chinese population. **PLoS One**, v. 7, n. 5, p. e37054, 2012.

SHOUKRY, N.H. et al. Cell-mediated immunity and the outcome of hepatitis C virus infection. **Annual review of microbiology**, v. 58, n. p. 391-424, 2004.

SIDWELL, R.W. et al. Broad-spectrum antiviral activity of Virazole: 1-beta-D-ribofuranosyl-1,2,4-triazole-3-carboxamide. **Science**, v. 177, n. 4050, p. 705-706, 1972.

SIMEN, B.B. et al. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. **J Infect Dis**, v. 199, n. 5, p. 693-701, 2009.

SIMMONDS, P. et al. Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. **J Gen Virol**, v. 75 (Pt 5), n. p. 1053-1061, 1994.

SIMMONDS, P. Genetic diversity and evolution of hepatitis C virus--15 years on. **J Gen Virol**, v. 85, n. Pt 11, p. 3173-3188, 2004.

SOLDAN, K. et al. Estimation of the risk of hepatitis B virus, hepatitis C virus and human immunodeficiency virus infectious donations entering the blood supply in England, 1993-2001. **Vox sanguinis**, v. 84, n. 4, p. 274-286, 2003.

SY, T. & JAMAL, M.M. Epidemiology of hepatitis C virus (HCV) infection. **International journal of medical sciences**, v. 3, n. 2, p. 41-46, 2006.

TAM, R.C. et al. Ribavirin polarizes human T cell responses towards a Type 1 cytokine profile. **J Hepatol**, v. 30, n. 3, p. 376-382, 1999.

TANAKA, T. et al. Structure of the 3' terminus of the hepatitis C virus genome. **J Virol**, v. 70, n. 5, p. 3307-3312, 1996.

TANAKA, Y. et al. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. **Nat Genet**, v. 41, n. 10, p. 1105-1109, 2009.

TAPPER, E.B. & AFDHAL, N.H. Is 3 the new 1: perspectives on virology, natural history and treatment for hepatitis C genotype 3. **J Viral Hepat**, v. 20, n. 10, p. 669-677, 2013.

TAYLOR, D.R. et al. Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. **Science**, v. 285, n. 5424, p. 107-110, 1999.

THOMAS, D.L. et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. **Nature**, v. 461, n. 7265, p. 798-801, 2009.

TORRES-PUENTE, M. et al. Hepatitis C virus and the controversial role of the interferon sensitivity determining region in the response to interferon treatment. **J Med Virol**, v. 80, n. 2, p. 247-253, 2008a.

TORRES-PUENTE, M. et al. Using evolutionary tools to refine the new hypervariable region 3 within the envelope 2 protein of hepatitis C virus. **Infect Genet Evol**, v. 8, n. 1, p. 74-82, 2008b.

TOYODA, H. et al. Quasispecies nature of hepatitis C virus and response to alpha interferon: significance as a predictor of direct response to interferon. **J Hepatol**, v. 26, n. 1, p. 6-13, 1997.

TREMBLING, P.M. et al. Treatment decisions and contemporary versus pending treatments for hepatitis C. **Nat Rev Gastroenterol Hepatol**, v. n. p. 2013.

TROESCH, M. et al. Study of a novel hypervariable region in hepatitis C virus (HCV) E2 envelope glycoprotein. **Virology**, v. 352, n. 2, p. 357-367, 2006.

UZE, G. & MONNERON, D. IL-28 and IL-29: newcomers to the interferon family. **Biochimie**, v. 89, n. 6-7, p. 729-734, 2007.

VELOSA, J. et al. Recomendações para a terapêutica da hepatite C. **GE J Port Gastrenterol**, v. 19, n. p. 6, 2012.

WANG, C. et al. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. **Genome Res**, v. 17, n. 8, p. 1195-1201, 2007.

WARTELLE-BLADOU, C. et al. Hepatitis C therapy in non-genotype 1 patients: the near future. **J Viral Hepat**, v. 19, n. 8, p. 525-536, 2012.

WASLEY, A. & ALTER, M.J. Epidemiology of hepatitis C: geographic differences and temporal trends. **Seminars in liver disease**, v. 20, n. 1, p. 1-16, 2000.

WEINER, A.J. et al. Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. **Virology**, v. 180, n. 2, p. 842-848, 1991.

WEINER, A.J. et al. Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. **Proc Natl Acad Sci U S A**, v. 89, n. 8, p. 3468-3472, 1992.

WITKOWSKI, J.T. et al. Design, synthesis, and broad spectrum antiviral activity of 1- -D-ribofuranosyl-1,2,4-triazole-3-carboxamide and related nucleosides. **J Med Chem**, v. 15, n. 11, p. 1150-1154, 1972.

YAGNIK, A.T. et al. A model for the hepatitis C virus envelope glycoprotein E2. **Proteins**, v. 40, n. 3, p. 355-366, 2000.

ZHOU, S. et al. The effect of ribavirin and IMPDH inhibitors on hepatitis C virus subgenomic replicon RNA. **Virology**, v. 310, n. 2, p. 333-342, 2003.

ZHOU, Y.H. et al. Monoclonal antibodies to the hypervariable region 1 of hepatitis C virus capture virus and inhibit virus adsorption to susceptible cells in vitro. **Virology**, v. 269, n. 2, p. 276-283, 2000.

Capítulo 2

Artigos Científicos

Hepatitis C virus genotype 3 and outcome to IFN therapy

Abstract. Hepatitis C is a major public health problem. New HCV antiviral drugs were released on the market in 2010; however, except for genotype 1, the most common therapy currently used is still based on Interferon (IFN) and Ribavirin. At the present, genotype 3 is the one with the highest rate of treatment failure. Viral genome variability is one of the factors that leads to therapy failure. HCV presents high mutation rates during the replication course, originating intra-host variants called quasispecies. The hypervariable region 1 (HVR1) from the envelope protein presents quasispecies and may be related to IFN therapy resistance. Regarding this issue, we determined HVR1 quasispecies from 14 patients infected with HCV genotype 3 using ultra-deep pyrosequencing (UDPS). We obtained 64,400 HVR1 sequences from serum samples of pre-therapy patients. Using the pipeline developed by DVH-CDC group, 27,398 sequences with high quality were filtered. Genetic distance and Shannon entropy values were not related to therapy outcome. These sequences were analyzed using median-joining networks and Bayesian population structural analysis. These analyses identified samples with different structures, from high conserved (one sub-population) to high stratified ones (6 sub-populations). Networks analysis also confirmed this result. Mutations that are exclusive for a type of response to therapy were identified along HVR1. These results shall contribute to the understanding of HCV quasispecies dynamics and therapy and how a high resolution tool as UDPS is essential to it.

INTRODUCTION

Hepatitis C disease is known to present high probability of persistence and characteristic progression, which includes cirrhosis and hepatocarcinogenesis (3, 12). It is estimated that 300 million people are infected worldwide and 3-4 million new cases are reported per year (6, 16, 35).

The etiological agent of hepatitis C is the hepatitis C virus (HCV) (Flaviviridae family member, Hepacivirus genus) (33). The genome is composed of a single strand positive RNA. During cell infection, it is translated into a polyprotein and cleaved in structural and nonstructural proteins (23, 25, 27). HCV is classified into 7 genotypes and more than 80 subtypes (20, 33). In addition, HCV circulates as a population of closely related variants, known as quasispecies, in infected patints (9, 10). This high variability is due to a combination of factors, including lack of correction activity by RNA-dependent RNA polymerase (RdRp), high viral replication rate and possible events of recombination (5, 26). Mutation rates are different along HCV genome: some regions show high conservation (ex. 5' and 3'UTR) and other regions, high variability (ex. E2) (30, 37, 41).

Viral persistence and escape of immune system may be related to the most variable region in HCV genome, the hypervariabilty region 1 (HVR1) in aminoterminal of E2 protein (18, 41). HVR1 is a 27 amino acid sequence in E2 surface not essential for virus replication; but it seems to play an important role in viral infectivity and as immune "decoy" (11, 41). Specific features of this region may also be useful to predict of Interferon (IFN) based therapy (13, 29, 39). For genotype 1, direct activity antiviral (DAA) drugs were recently implemented as World Health Organization indicated therapy, showing an efficiency of 70 to 90% in patients (7, 22). However, there are still 50-70 millions of people infected with other HCV genotypes, who drag clinicians and researches attention, since they are still a major public health problem. In this context, genotype 3 presents the lower sustained virological rate to therapy and is distributed worldwide. HCV-3 is prevalent in South Asia, Europe, Oceania, Canada and in some countries of South America. Around 30% of the patients infected with genotype 3 are resistant to IFN based therapy. (15, 38, 40). Regarding this issue, this study applied the ultra-deep pyrosequencing for HVR1 in 14 baseline samples, collected from patients infected with HCV genotype 3. Analysis of HVR1 revealed differential features in composition of patients with different response to quasispecies, indicating that HVR1 composition before the IFN therapy is important to treatment outcome.

MATERIAL AND METHODS

Patients and samples. Baseline sera samples from 14 patients infected with Genotype 3, treated by combined therapy IFN+Ribavirin for 24 weeks were used. Response to therapy was determined by detection of viral RNA,

classifying patients in three response groups: sustained virological responders (SVR), if HCV-RNA was detected only before treatment administration; non-responders (NR), if it was detected before, during and after end of treatment; and at last relapses (ETR), if viral RNA was not detected during, but became detectable in some time point after the end of therapy. Patients with clinical history of alcohol high consumption (>30g/day) or co-infection with other virus which causes liver damage (ex. HBV) or immune deficiency (ex. HIV) were excluded. All patients were enrolled at Sao Jose do Rio Preto city Blood Center and consented participation, according to Brazilian ethical rules in human researches.

HCV-HVR1 amplification and consensus sequencing. RNA from 100 µL sera was extracted by MagnaPureLC kit and MagnaPure Automated System (Roche, Inc., Indianapolis, IN). Eluted RNA was submitted to cDNA synthesis by cDNA SuperScript Vilo (Invitrogen, by Life Technologies). Amplification of HVR1 by PCR and Nested-PCR was performed using PerfeCta SYBR Green FastMix (Quanta Biosciences) LightCycler 480 instrument (Roche,Inc.) with primers and protocol previously described (31). Positive amplification samples were selected by melting curve analysis. Consensus sequences were obtained by BigDye v3.1 chemistry sequencing kit (Applied Biosystems, Foster City, CA) in automated sequencer (3130xl genetic analyzer; Applied Biosystems).

Ultra-deep pyrossequencing. First PCR products from HVR1 successful amplified samples were used for a new amplification round using primers from Nested-PCR modified with MID tags and adaptors. Fourteen set of primers were used and their sequence are described in Table S1 and S2. PCR products were purified by E-Gel® SizeSelect™ (Invitrogen, by Life Technologies). These purified products quality and quantity were verified in 2100 Bioanalyzer (Agilent DNA 1000 Kit, Agilent Technologies). Based on this quantification result, samples were diluted and pooled together in equimolar ratios. This mixture was submitted to pyrosequencig according to Roche/454 GS Junior (454 Sequencing, Roche) protocol for amplificon, described by manufacturer.

Data correction/filtering and analysis. Sequences obtained from pyrosequencing were corrected and filtered according to protocol based on k-mers error correction and empirical frequency threshold (34). Mega 5.2 software (36) was used for haplotype alignment and singletons were removed for further analysis. Mega was also used for genetic distance, dn/ds calculations and substitution model definition. Shannon-Entropy and mutation profile was analysed using Entropy-One (19) and Highlighter (42) tools at HIV databases from Los Alamos National Laboratory (http://www.hiv.lanl.gov/content/sequence/HIV/HIVTools.html). Maximum-Likelihood phylogenetic tree was built using PhyML (14) with general time reversible (GTR) model, determined by FindModel at Mega 5.2 and 100 bootstrap replicates. Median-joining networks (MJN) were applied to find minimum spanning phylogeny tree using Network 4.0 (4). Population structure analysis by Bayesian approaches

was performed by BAPS (4, 8) for the most 50 frequent haplotypes. Statistical hypothesis tests were calculated using Minitab 16® (Quality. Analysis. Results.®, Minitab, Inc.).

RESULTS

From the fourteen samples, 8 (57%) were male and 6 (43%) female of 28-67 years old (mean of 52.43 years old). Based on therapy outcome, 6 patients were responders, 5 were non-responders and 3 relapses. Viral load varied from 5.76-6.78 log10copies/mL (Table 1). No statistical significance was found between gender (chi-square test) or viral load (ANOVA with Turkey and Fisher post hoc test) and response to therapy. ANOVA test showed that Non-responder group is older than Responders (p=0.02).

Successful PCR amplification resulted in a fragment of approximately 400pb, confirmed by consensus sequencing. Pirosequencing from the baseline samples provided 64,400 HVR1 sequences. Error correction and filtering resulted in 27,398 sequences. A summary of the number of sequences before and after pipeline processing by sample is showed in Figure 1. The absolute number is detailed in Table S3. Pearson correlation test indicated that there was no statistical relationship between the number of sequences extracted from the machine and number of haplotypes or calculated genetic diversity, indicating that different numbers of sequences of HVR1 in each sample did not influence the further analysis of quasispecies. Total number of haplotypes varied from 1 (P31) to 568 (P28), with no relation to therapy outcome verified by statistical test (ANOVA with Turkey and Fisher post hoc test).

Pairwise and overall genetic distance were calculated. Overall genetic distance is showed in Figure 2. Responder patients showed the highest (P28=0.066) and the lowest (P31=0) distance value between the haplotypes.

Shannon-Entropy and Highlighter plots (at Supplementary Material) showed that mutations were concentrated around nucleotide positions 190-265. Maximum Shannon-Entropy value was higher in Responders (R) (0.976 at pos 244) and NonResponders (NR) (0.981 at pos 223) and lower in Relapses (RFT) (0.687 at pos 191). Some mutation positions were present only in one response group but not in others. Summary of these mutations are in Table 2. Also, rarely mutations varied from more than one type of nucleotide in the same patient (pos 244 and 263). Excepting P31, who presented only one haplotype, Responder patients showed high frequency of mutations, but lower median Entropy value along sequences than Non-responders/Relapses patients (Figure 3). Statistical test (Kruskall-Wallis) indicated significant difference between these values (p=0.035).

Maximum-Likelihood tree (Figure 4) showed that HVR1 sequences were clustered in a monophyletic branch by patient, but not by type of response. Detailed tree per patient is showed is added at Supplementary Material.

Bayesian analysis of population structure (BAPS) and Median-joining networks distinguished homogeneous populations with only one cluster at patients P31 and P51 (subpopulation) and quasispecies populations of even six subpopulations in patient P62 (at Supplementary Material). There were no relations between heterogeneity and response to therapy. However, focusing on the number of mutations in each pool (Table 3), it is possible to observe that NR presents lower number of mutations than R and RFT.

DISCUSSION

HVR1 pretreatment variants are suggested by many studies to be related to IFN therapy, mainly in genotype 1(13). Nevertheless, even with all these efforts, it remains inconclusive the correlation with genetic heterogeneity in HVR1 and therapy response. There are studies reporting high diversity of HVR1 quasispecies and failure to respond (25, 27, 38,42). However, greater rates of non-synonymous substitutions than synonymous are observed in responders, indicating that HVR1 reflects immune response pressure involved in clearance during therapy (34). HVR1 domain may also be related to infection persistence (7).

Although recommended therapy for genotype 1 is based on DAA antivirals, genotype 3 is treated with IFN (38). Available data on relation between HCV quasispecies and IFN are derived mostly from genotype 1 studies(1, 2, 17). However, it is suggested that the dynamics of quasispecies is different among genotypes and may be a partial explanation to differences in response rates (21, 28, 38).

In the presented study, we detailed quasispecies from HCV genotype 3 based on HVR1 baseline sequences of patients with different therapy outcomes. These patients showed no significant differences of viral loads at this point.

Some studies pointed to a higher diversity in HVR1 quasispecies related to nonresponse patients (29, 43). However, like other studies (24, 32), our study did not report this correlation. Although overall average of Shannon-Entropy between the three groups were lower in R group, patient with lowest and highest diversity were from this same group (Figure 2 and 4). This result suggests that, for genotype 3, quasispecies in responders have a greater range of diversity than genotype 1.

Some specific positions of mutations were present in only one response group (Table 2). Point mutations were analyzed in previous studies (Gaudy; Kim, 2012), which were separated by response group. Although it is indicative of response type specific mutation, it could not be compromised, since host factors also play an important role in directing mutations at HVR1 region.

Subpopulation and Median-Joining networks indicated that HVR1 variants before IFN therapy can be organized in several communities in most of the patients, indicating heterogeneity of quasispecies populations. Only one R patient (P31) and one NR patient (P51) showed a homogeneous characterized variants, defined as one structure of population. This indicates that the structure of population does not reflect on response, however it may be a hint to how host immune system is working on HCV before IFN activity: an homogeneous population indicates high negative pressure on quasispecies, targeting a broad spectrum of pool of variants; on other hand, high diversity suggests that immune system is working on a smaller pool of quasispecies, as many variants are escaping (34). Interestingly, as observed previously (13) it is not response related.

In conclusion, the present results indicate that HVR1 structures may acquire different features patients before IFN administration and does not reflect outcome to therapy. However, it is important to highlight that there are indications of divergence comparing genotype 1, which does not allow the same interpretation in both cases. Considering this fact, genotype 3 quasispecies analysis during infection course is important to hepatitis C virus dynamics understanding.

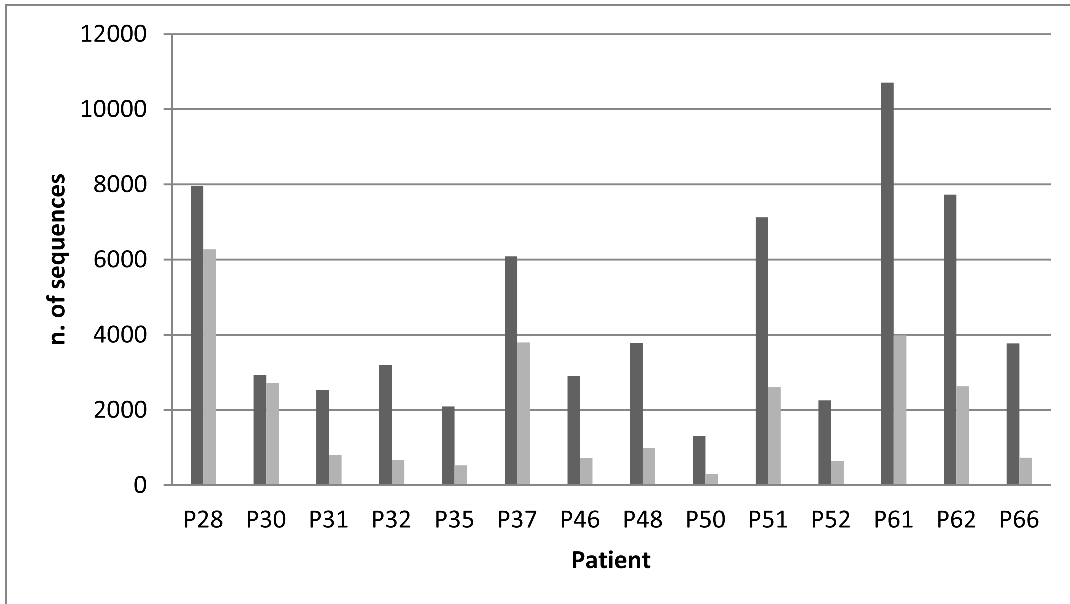


Fig. 1. Number of sequences before (dark grey) and after (light gray) pipeline processing.

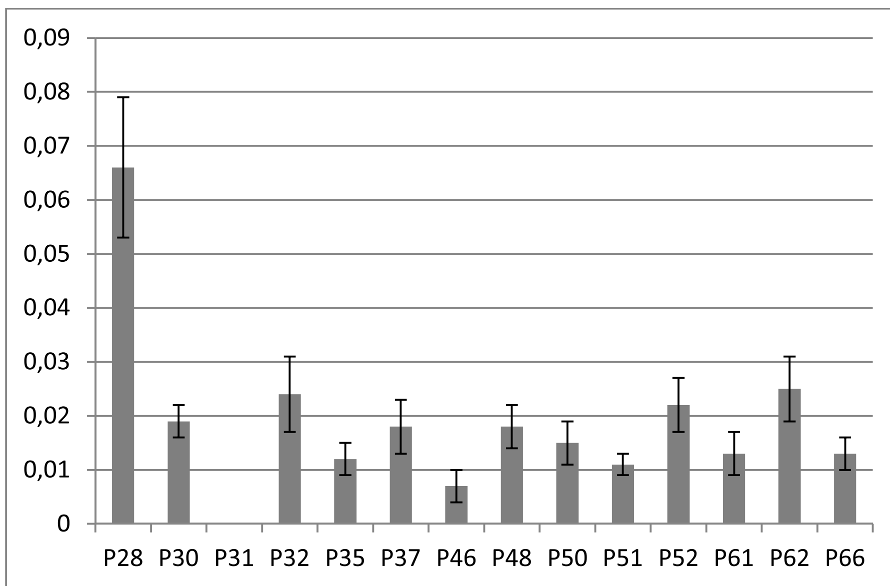


Fig. 2. Overall mean of genetic distance in each patient baseline sample.

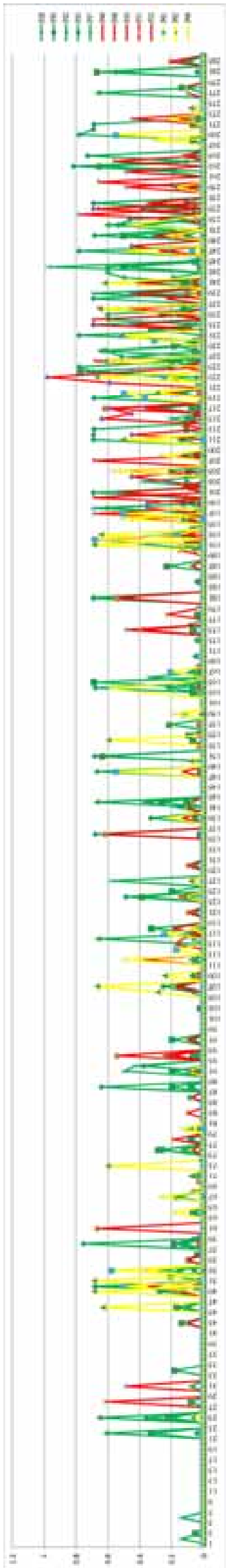


Figure 3. Shannon-Entropy plot colored by response, each line correspond to one patient. X-axis – nt position; y-axis – Entropy value. Green – Responders; Red – NonResponders and Yellow – Relapses.

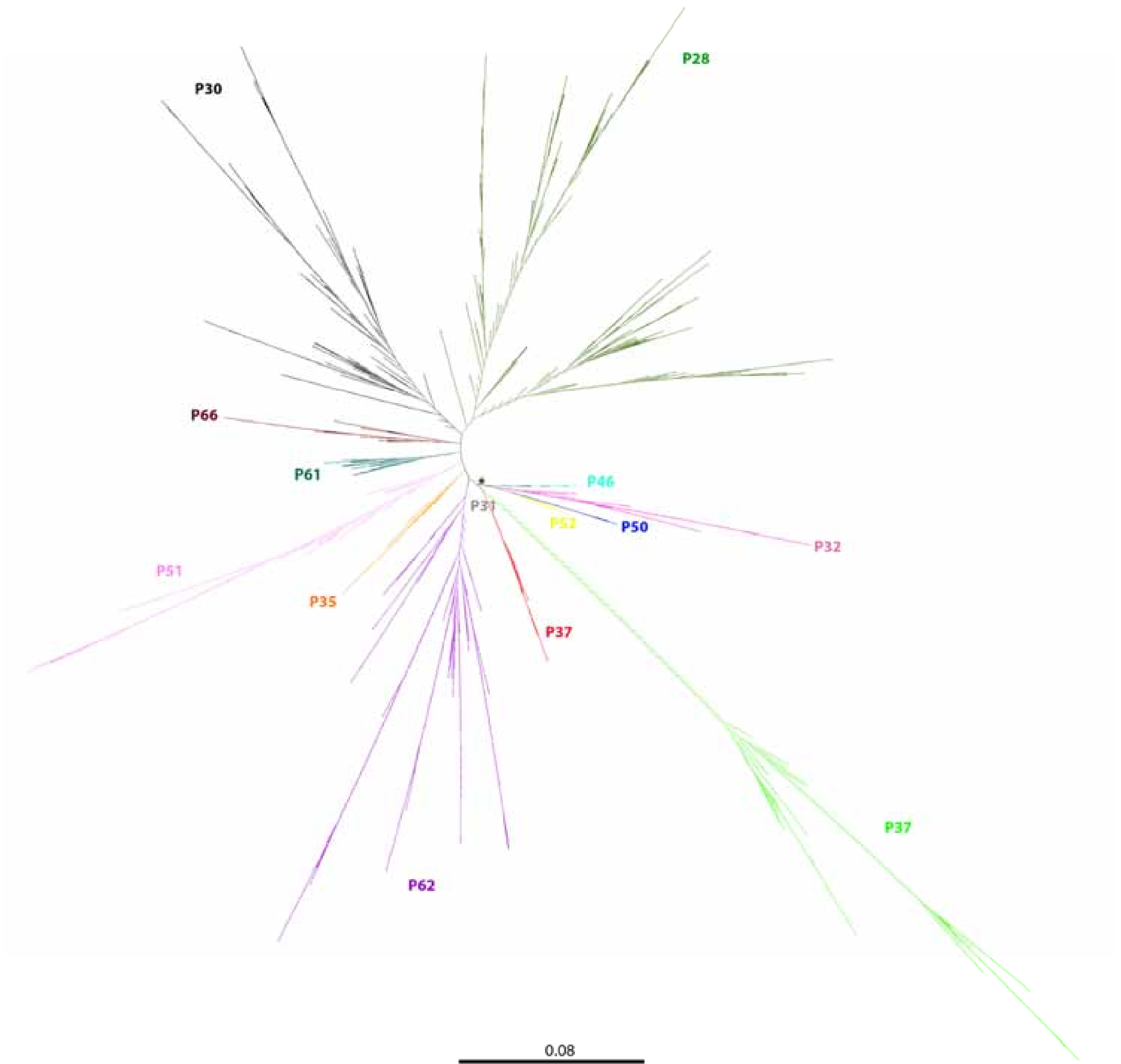


Figure 4. ML phylogenetic tree, colored by patient. "*" shows node with bootstrap >70%.

Table 1. Patients codes and characteristics.

Patient number	Gender*	Age (years)	Viral load (log_{10} copies/mL)	Response
P28	F	54	6,78	Responder
P30	M	54	6,68	Responder
P31	M	58	NA**	Responder
P32	M	60	NA	Responder
P35	F	65	NA	Responder
P37	M	67	NA	Responder
P46	M	39	NA	Non-responder
P48	M	51	6,28	Non-responder
P50	M	48	6,5	Non-responder
P51	F	46	6,5	Non-responder
P52	F	28	5,76	Non-responder
P61	M	41	NA	Relapse
P62	F	66	6,42	Relapse
P66	F	57	5,79	Relapse

*F – Female, M – Male **NA – not available

Table 2. Mutations found in each position, exclusive from each response group.

Position	Mutation	Response group	Position	Mutation	Response group
2	G→T	R(P37)	166	G→A	R(P30)
3	A→G	R(P30)	168	T→C	NR(P51)
6	T→G	R(P37)	170	A→G	R(P28)
22	C→T T→A/T	R(P28,P235) R(P32)	176	T→G	R(P28)
34	T→A	R(P30)	179	G→A	NR (P51)
52	C→T	R(P30)	184	C→T	R (P28, P30)
56	C→T	R(P30)	194	A→G	R (P30)
59	T→C C→T	R(P28) R(P30)	207	C→G	NR (P50)
61	G→A	NR(P46)	213	C→T	R (P30)
70	G→T	NR(P51)	218	G→A	R (P30)
73	G→A	RFT(P62)	227	G→A A→G	R (P28) R(P37)
74	T→G	RFT(P62)	228	C→T T→C	R (P28) R(P30)
79	C→T	R(P30)	229	C→T	R (P28, P30)
83	A→G	NR(P51)	240	T→C	R (P28)
88	C→T G→A	R(P28,P30) R(P35)	242	A→G	R (P28, P32)
92	G→A	P(28,P37)	243	A→G	R (P32)
103	C→T	R(P30)	244	C→T G→C A→T T→G/C	R (P30) R(P32) R(P35) R(P37)
114	G→A	RFT(P61)	245	C→T T→C	R(P30) R(P37)
115	C→T	NR(P51)	248	T→C	NR (P52)
125	T→C	R(P35)	260	T→G	NR (P50)
140	A→G	R(P30)	263	A→C/G C→A	R(P28) R(P30)
151	A→G	R(P28, P30)	270	A→C	R(P28)
154	T→C	RFT(P62)	271	C→A A→G	R(P28) R(P30)
159	T→C	RFT(P61, P62, P66)	281	G→A	R(P28, P30, P32)

R – Responders; NR – NonResponders and RFT – Relapses.

Table 3. Networks and BAPS analysis.

Patient number	N. of clusters	N. of mutations steps
P28	3	57
P30	2	24
P31	1	-
P32	3	48
P35	3	32
P37	4	41
P46	4	9
P48	4	35
P50	3	20
P51	1	36
P52	3	28
P61	4	20
P62	6	57
P66	3	33

ACKNOWLEGMENTS

Financial support by FAPESP (project n. 2010/00976-0) and CAPES (project n. BEX 3780/10-4).

REFERENCES

1. Abbasi, A., A. R. Bhutto, N. Butt, and S. M. Munir. 2011. Efficacy of interferon-ribavirin therapy in chronic hepatitis C patients. J Coll Physicians Surg Pak 21:276-279.

2. Afzal, S., M. Idrees, M. Ali, M. Ilyas, A. Hussain, M. Akram, S. Butt, S. Saleem, I. U. Rehman, L. Ali, and M. Shahid. 2011. Envelope 2 protein phosphorylation sites S75 & 277 of hepatitis C virus genotype 1a and interferon resistance: a sequence alignment approach. Virol J 8:71.

3. Alter, H. J., R. H. Purcell, J. W. Shih, J. C. Melpolder, M. Houghton, Q. L. Choo, and G. Kuo. 1989. Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. N Engl J Med 321:1494-1500.

4. Bandelt, H. J., P. Forster, and A. Rohl. 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular biology and evolution 16:37-48.

5. Caraballo Cortes, K., O. Zagordi, T. Laskus, R. Ploski, I. Bukowska-Osko, A. Pawelczyk, H. Berak, and M. Radkowski. 2013. Ultradeep pyrosequencing of hepatitis C virus hypervariable region 1 in quasispecies analysis. BioMed research international 2013:626083.

6. Chambers, T. J., X. Fan, D. A. Droll, E. Hembrador, T. Slater, M. W. Nickells, L. B. Dustin, and A. M. Dibisceglie. 2005. Quasispecies heterogeneity within the E1/E2 region as a pretreatment variable during pegylated interferon therapy of chronic hepatitis C virus infection. Journal of virology 79:3071-3083.

7. Conteduca, V., D. Sansonno, S. Russi, F. Pavone, and F. Dammacco. 2013. Therapy of chronic hepatitis C virus infection in the era of direct-acting and host-targeting antiviral agents. J Infect.

8. Corander, J., P. Marttinen, J. Siren, and J. Tang. 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinformatics 9:539.

9. Domingo, E., V. Martin, C. Perales, A. Grande-Perez, J. Garcia-Arriaza, and A. Arias. 2006. Viruses as quasispecies: biological implications. Current topics in microbiology and immunology 299:51-82.

10. Farci, P. 2011. New insights into the HCV quasispecies and compartmentalization. Semin Liver Dis 31:356-374.

11. Forns, X., R. Thimme, S. Govindarajan, S. U. Emerson, R. H. Purcell, F. V. Chisari, and J. Bukh. 2000. Hepatitis C virus lacking the hypervariable region 1 of the second envelope protein is infectious and causes acute resolving or persistent infection in chimpanzees. Proc Natl Acad Sci U S A 97:13318-13323.

12. Gale, M., Jr., and E. M. Foy. 2005. Evasion of intracellular host defence by hepatitis C virus. Nature 436:939-945.

13. Gaudy, C., A. Moreau, P. Veillon, S. Temoin, F. Lunel, and A. Goudeau. 2003. Significance of pretreatment analysis of hepatitis C virus genotype 1b hypervariable region 1 sequences to predict antiviral outcome. J Clin Microbiol 41:3615-3622.

14. Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology 59:307-321.

15. Hajarizadeh, B., J. Grebely, and G. J. Dore. 2013. Epidemiology and natural history of HCV infection. Nat Rev Gastroenterol Hepatol 10:553-562.

16. Huang, R. H., and K. Q. Hu. 2006. A practical approach to managing patients with HCV infection. International journal of medical sciences 3:63-68.

17. Jardim, A. C., L. H. Yamasaki, A. T. de Queiroz, C. Bittar, J. R. Pinho, C. M. Carareto, P. Rahal, and I. M. Mello. 2009. Quasispecies of hepatitis C virus genotype 1 and treatment outcome with peginterferon and ribavirin. Infect Genet Evol 9:689-698.

18. Kato, N., Y. Ootsuyama, S. Ohkoshi, T. Nakazawa, H. Sekiya, M. Hijikata, and K. Shimotohno. 1992. Characterization of hypervariable regions in the putative envelope protein of hepatitis C virus. Biochem Biophys Res Commun 189:119-127.

19. Keele, B. F., E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham, M. G. Salazar, C. Sun, T. Grayson, S. Wang, H. Li, X. Wei, C. Jiang, J. L. Kirchherr, F. Gao, J. A. Anderson, L. H. Ping, R. Swanstrom, G. D. Tomaras, W. A. Blattner, P. A. Goepfert, J. M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, M. S. Cohen, D. C. Montefiori, B. F. Haynes, B. Gaschen, G. S. Athreya, H. Y. Lee, N. Wood, C. Seoighe, A. S. Perelson, T. Bhattacharya, B. T. Korber, B. H. Hahn, and G. M. Shaw. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A 105:7552-7557.

20. Kuiken, C., and P. Simmonds. 2009. Nomenclature and numbering of the hepatitis C virus. Methods Mol Biol 510:33-53.

21. Kumthip, K., C. Pantip, P. Chusri, S. Thongsawat, A. O'Brien, K. E. Nelson, and N. Maneekarn. 2011. Correlation between mutations in the core and NS5A genes of hepatitis C virus genotypes 1a, 1b, 3a, 3b, 6f and the response to pegylated interferon and ribavirin combination therapy. J Viral Hepat 18:e117-125.

22. Lawitz, E., and E. J. Gane. 2013. Sofosbuvir for previously untreated chronic hepatitis C infection. N Engl J Med 369:678-679.

23. Liang, Y., H. Ye, C. B. Kang, and H. S. Yoon. 2007. Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded. Biochemistry 46:11550-11558.

24. Lopez-Labrador, F. X., S. Ampurdanes, M. Gimenez-Barcons, M. Guilera, J. Costa, M. T. Jimenez de Anta, J. M. Sanchez-Tapias, J. Rodes, and J. C. Saiz. 1999. Relationship of the genomic complexity of hepatitis C virus with liver disease severity and response to interferon in patients with chronic HCV genotype 1b infection [correction of interferon]. Hepatology 29:897-903.

25. Lyra, A. C., X. Fan, and A. M. Di Bisceglie. 2004. Molecular biology and clinical implication of hepatitis C virus. Braz J Med Biol Res 37:691-695.

26. Martell, M., J. I. Esteban, J. Quer, J. Genesca, A. Weiner, R. Esteban, J. Guardia, and J. Gomez. 1992. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. J Virol 66:3225-3229.

27. Moradpour, D., A. Cerny, M. H. Heim, and H. E. Blum. 2001. Hepatitis C: an update. Swiss Med Wkly 131:291-298.

28. Neumann, A. U., N. P. Lam, H. Dahari, M. Davidian, T. E. Wiley, B. P. Mika, A. S. Perelson, and T. J. Layden. 2000. Differences in viral dynamics between genotypes 1 and 2 of hepatitis C virus. J Infect Dis 182:28-35.

29. Okada, S., Y. Akahane, H. Suzuki, H. Okamoto, and S. Mishiro. 1992. The degree of variability in the amino terminal region of the E2/NS1 protein of hepatitis C virus correlates with responsiveness to interferon therapy in viremic patients. Hepatology 16:619-624.

30. Okamoto, H., S. Okada, Y. Sugiyama, S. Yotsumoto, T. Tanaka, H. Yoshizawa, F. Tsuda, Y. Miyakawa, and M. Mayumi. 1990. The 5'-terminal sequence of the hepatitis C virus genome. The Japanese journal of experimental medicine 60:167-177.

31. Ramachandran, S., D. S. Campo, Z. E. Dimitrova, G. L. Xia, M. A. Purdy, and Y. E. Khudyakov. 2011. Temporal Variations in the Hepatitis C Virus Intra-Host Population during Chronic Infection. J Virol.

32. Sandres, K., M. Dubois, C. Pasquier, J. L. Payen, L. Alric, M. Duffaut, J. P. Vinel, J. P. Pascal, J. Puel, and J. Izopet. 2000. Genetic heterogeneity of hypervariable region 1 of the hepatitis C virus (HCV) genome and sensitivity of HCV to alpha interferon therapy. J Virol 74:661-668.

33. Simmonds, P. 2004. Genetic diversity and evolution of hepatitis C virus--15 years on. J Gen Virol 85:3173-3188.

34. Skums, P., Z. Dimitrova, D. S. Campo, G. Vaughan, L. Rossi, J. C. Forbi, J. Yokosawa, A. Zelikovsky, and Y. Khudyakov. 2012. Efficient error correction for next-generation sequencing of viral amplicons. BMC Bioinformatics 13 Suppl 10:S6.

35. Sy, T., and M. M. Jamal. 2006. Epidemiology of hepatitis C virus (HCV) infection. International journal of medical sciences 3:41-46.

36. Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution 28:2731-2739.

37. Tanaka, T., N. Kato, M. J. Cho, K. Sugiyama, and K. Shimotohno. 1996. Structure of the 3' terminus of the hepatitis C virus genome. J Virol 70:3307-3312.

38. Tapper, E. B., and N. H. Afdhal. 2013. Is 3 the new 1: perspectives on virology, natural history and treatment for hepatitis C genotype 3. J Viral Hepat 20:669-677.

39. Toyoda, H., T. Kumada, S. Nakano, I. Takeda, K. Sugiyama, T. Osada, S. Kiriyama, Y. Sone, M. Kinoshita, and T. Hadama. 1997. Quasispecies nature of hepatitis C virus and response to alpha interferon: significance as a predictor of direct response to interferon. J Hepatol 26:6-13.

40. Wartelle-Bladou, C., G. Le Folgoc, M. Bourliere, and L. Lecomte. 2012. Hepatitis C therapy in non-genotype 1 patients: the near future. J Viral Hepat 19:525-536.

41. Weiner, A. J., H. M. Geysen, C. Christopherson, J. E. Hall, T. J. Mason, G. Saracco, F. Bonino, K. Crawford, C. D. Marion, K. A. Crawford, and et al. 1992. Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. Proc Natl Acad Sci U S A 89:3468-3472.

42. Yusim, K., C. Kesmir, B. Gaschen, M. M. Addo, M. Altfeld, S. Brunak, A. Chigaev, V. Detours, and B. T. Korber. 2002. Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. J Virol 76:8757-8768.

43. Zhang, L., F. Han, D. Zhang, and X. G. Dou. 2012. Mutations in different regions of the genome of hepatitis C virus genotype 1b and association with response to interferon therapy. Int J Mol Med 30:1438-1442.

SUPPLEMENTARY MATERIAL

Table S1. Primers used for HVR1 amplification. (Red – key sequence for pyrosequencing; Blue – One of the 14 MID sequences in Table S2).

Primer	Sequence	Step used
HVR1-F1	5'-tggcttgggatatgatgatgaact-3'	PCR
HVR1-R1	5'-gcagtcctgttgatgtgcca-3'	PCR
HVR1-F2	5'-ggatatgatgatgaactggt-3'	NESTED
HVR1-R2	5'-ttgatgtgccagctgccgttggtgt-3'	NESTED
HVR1-MID"X"F	5'-cgtatcgcctccctcgcgccatcag-[mid]-ggatatgatgatgaactggt-3'	454 Sequencing
HVR1-MID"X"F	5'-ctatgcgccttgccagcccgctcag-[mid]-ttgatgtgccagctgccgttggtgt-3'	454 Sequencing

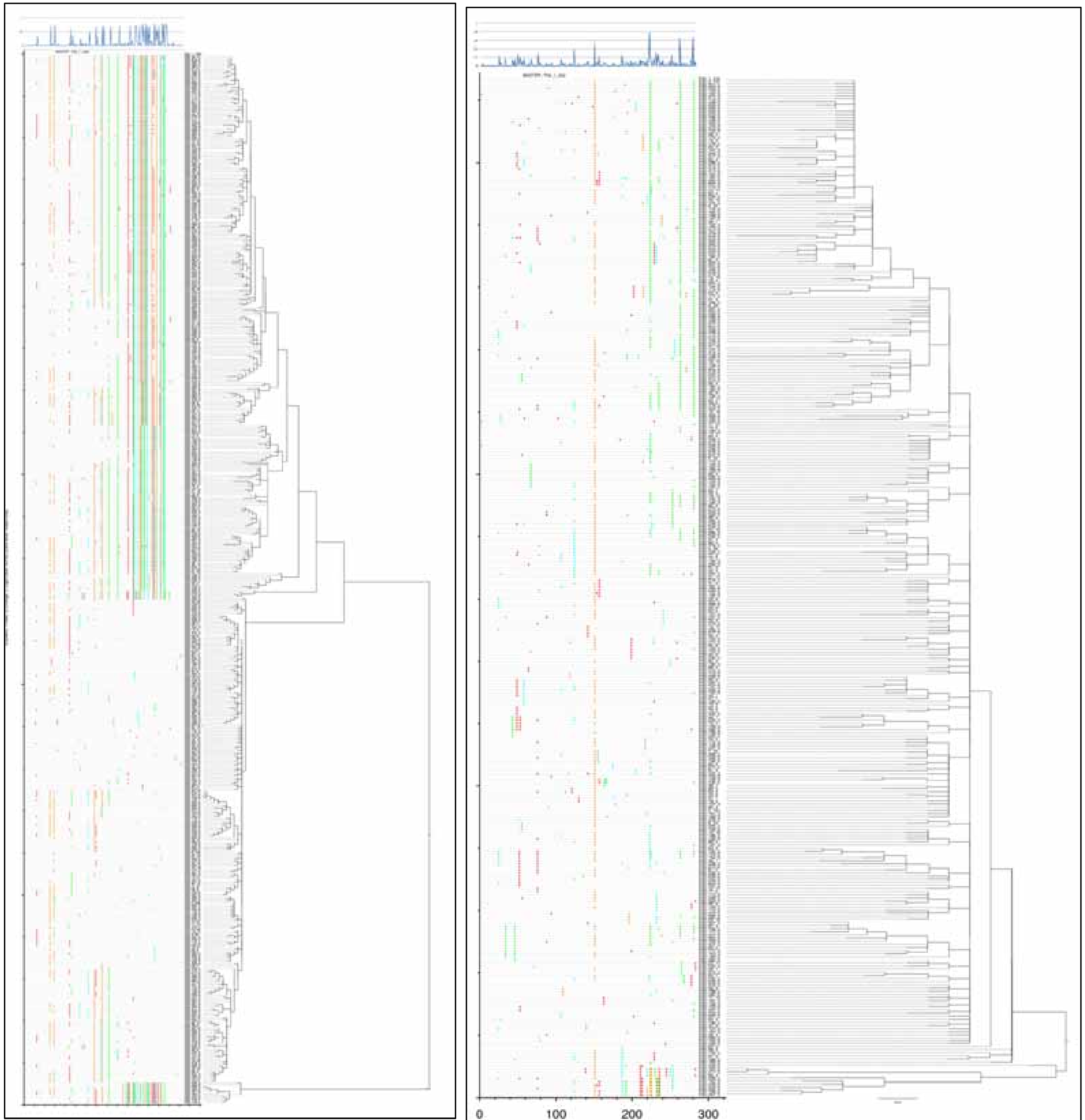
Table S2. MID adaptors sequences

Name	Sequence	Name	Sequence
MID1	ACGAGTGCGT	MID8	CTCGCGTGTC
MID2	ACGCTCGACA	MID9	TAGTATCAGC
MID3	AGACGCACTC	MID10	TCTCTATGCG
MID4	AGCACTGTAG	MID11	TGATACGTCT
MID5	ATCAGACACG	MID12	TACTGAGCTA
MID6	ATATCGCGAG	MID13	CATAGTAGTG
MID7	CGTGTCTCTA	MID14	CGAGAGATAC

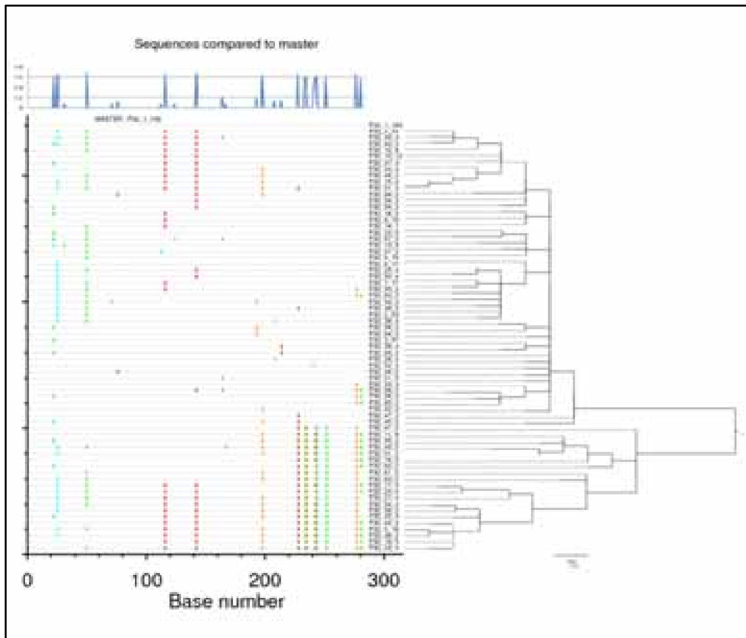
Table S3. Number of raw sequences, sequences after pipeline processing and nucleotide haplotypes per patient sample.

Patient	Raw sequences	Pipeline sequences	Nucleotide haplotypes
P28	7960	6271	568
P30	2929	2718	327
P31	2527	808	1
P32	3197	674	68
P35	2097	524	40
P37	6086	3797	226
P46	2905	723	13
P48	3786	990	42
P50	1305	295	16
P51	7130	2602	140
P52	2254	647	18
P61	10714	3985	76
P62	7734	2632	314
P66	3776	732	66

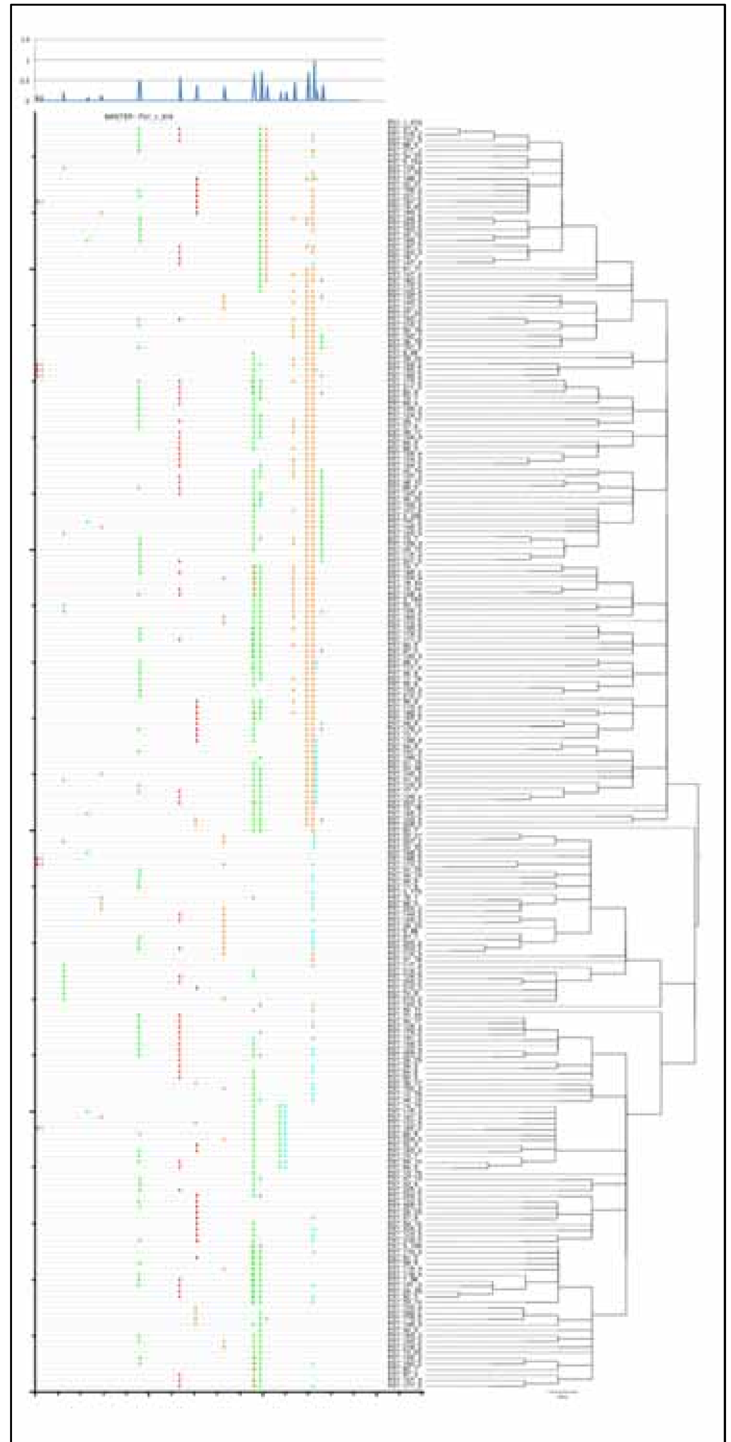
Figure S1. Highlighter and Shannon-Entropy plots for eah position. Master sequence used is the major haplotype. Order of the sequences is based on ML tree. Graph in blue – Shannon-entropy; color of each mutation: A-green, T-red, C – light blue, G – orange.



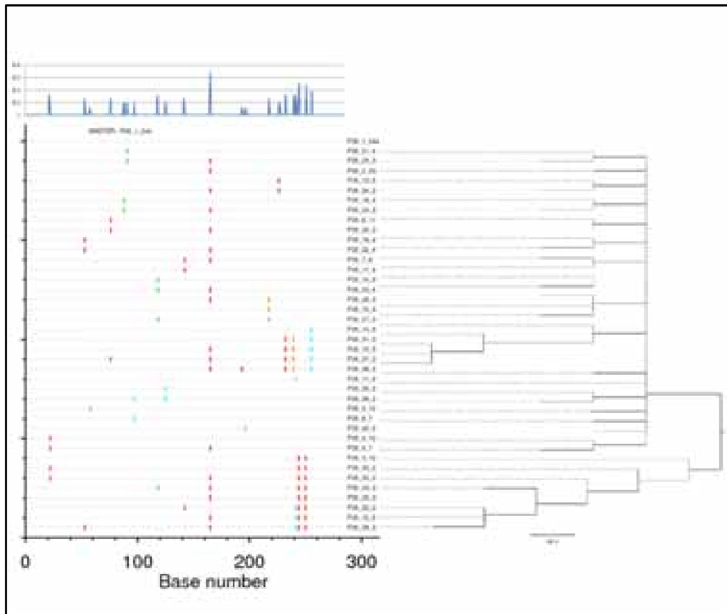
P32



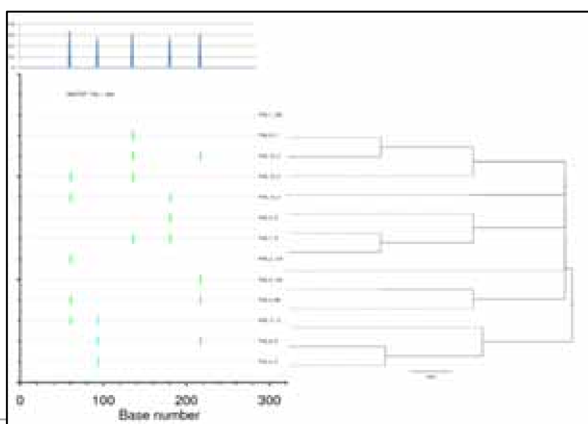
P37



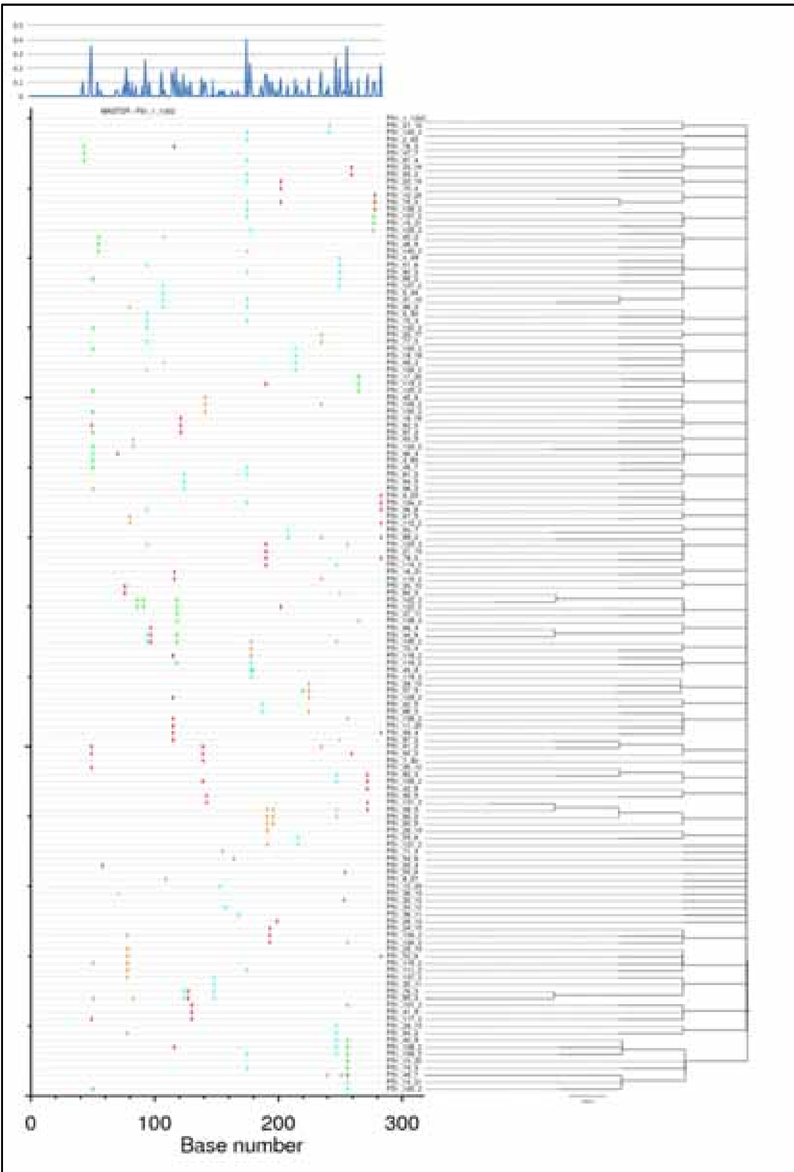
P35



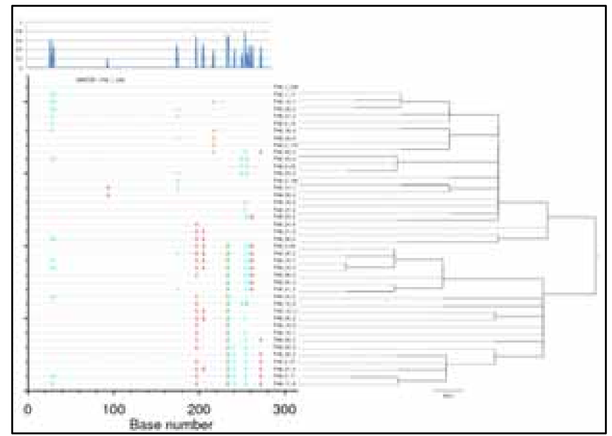
P46



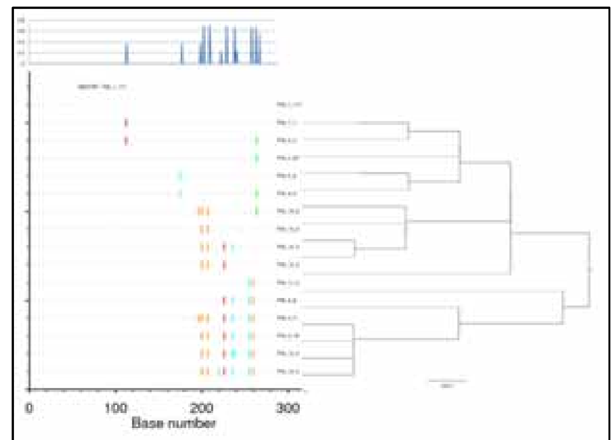
P51



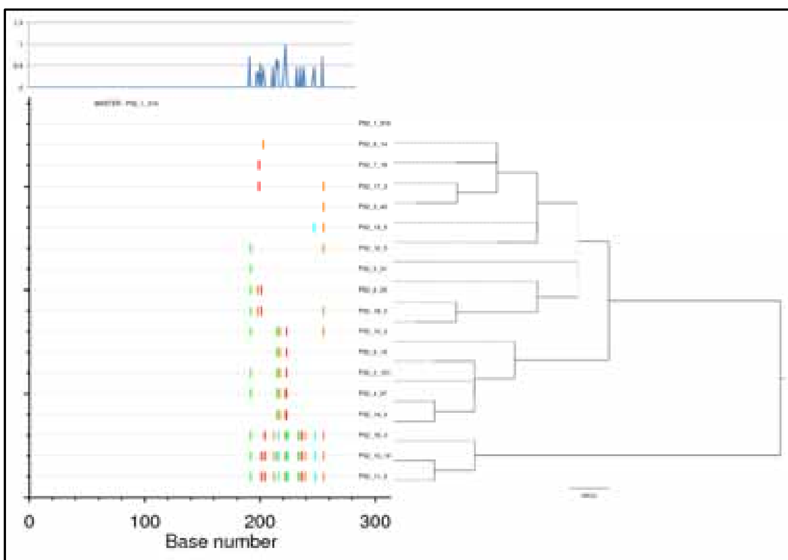
P48



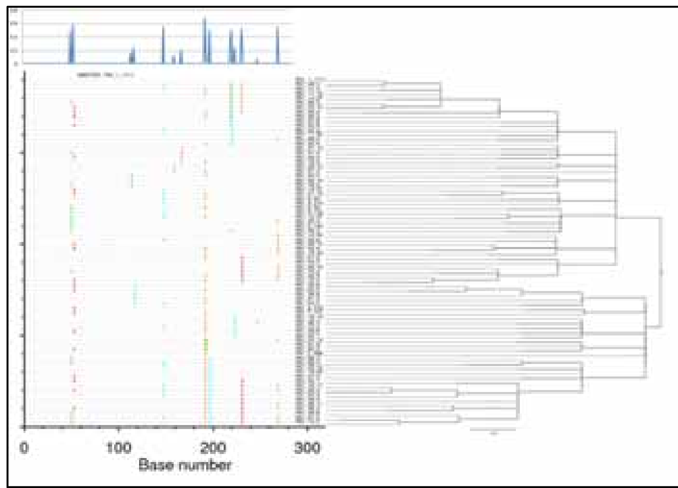
P50



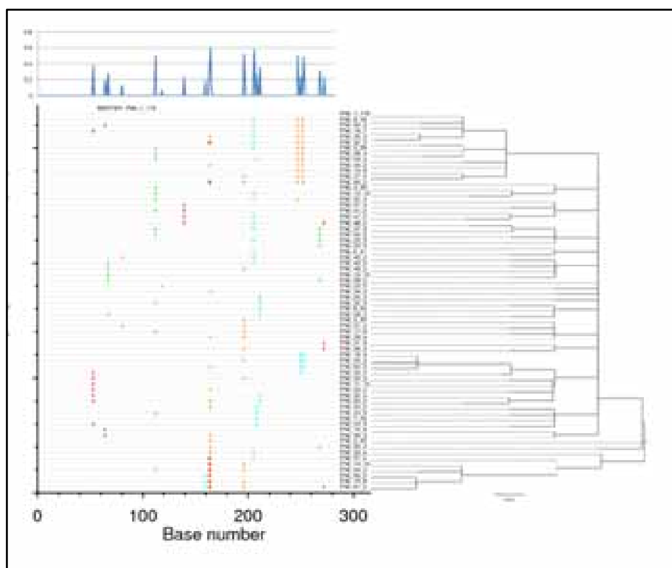
P52



P61



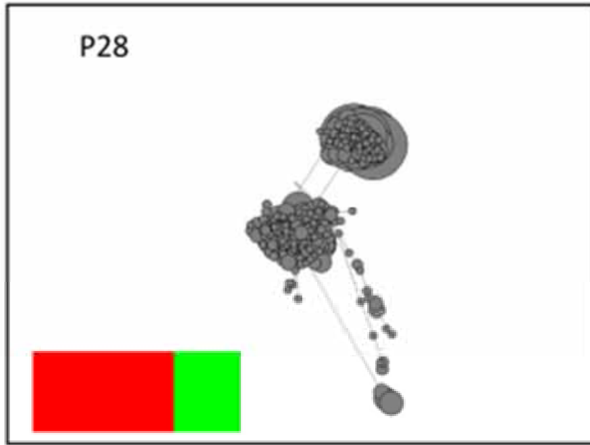
P66

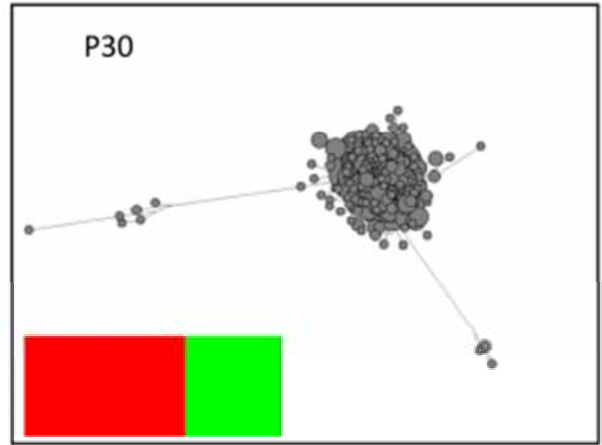


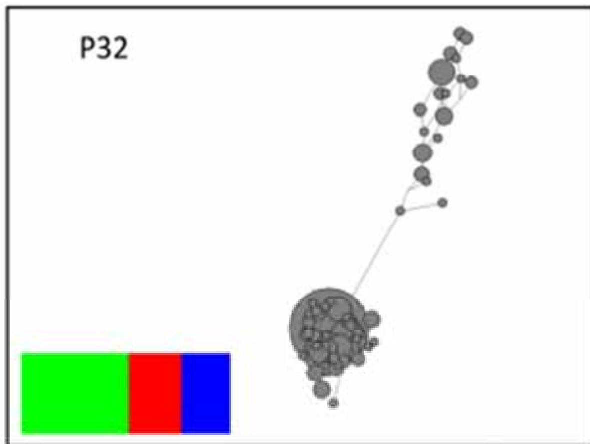
P62

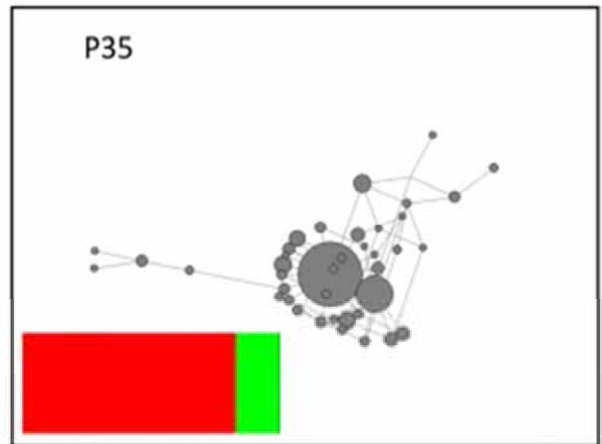
0
100
200
300
Base number

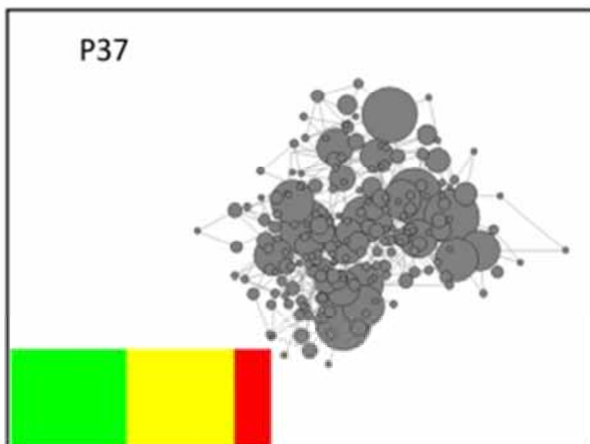
Fig S2. Median-joining networks and plot showing number and relative size of each subpopulations (right-down square).
Each circle represents a haplotype and size is according to it frequency. Each color in plot represents a subpopulation recognized and it relative size in total population.

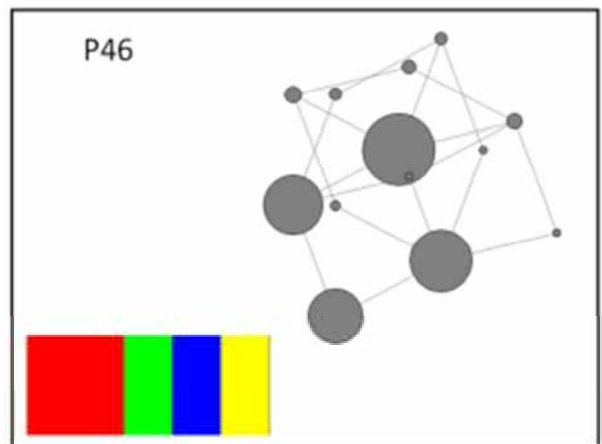


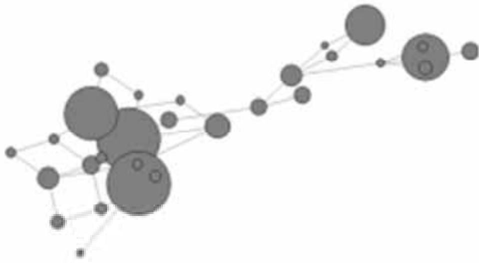
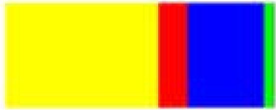
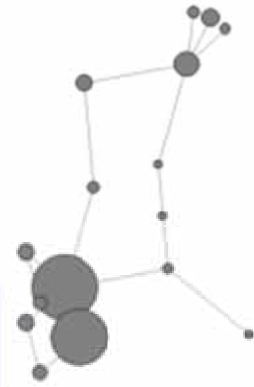
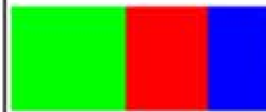
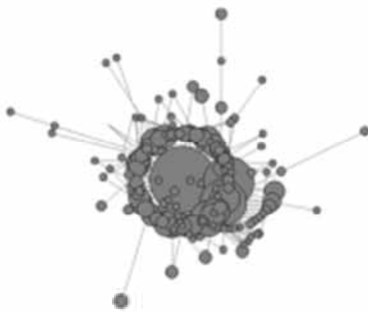
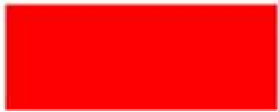
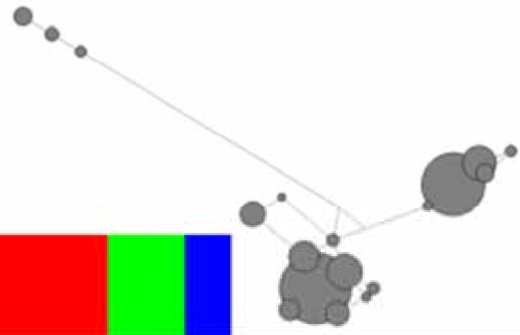
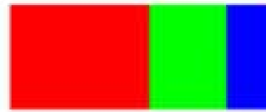
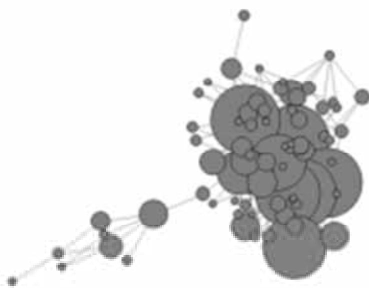
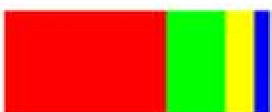
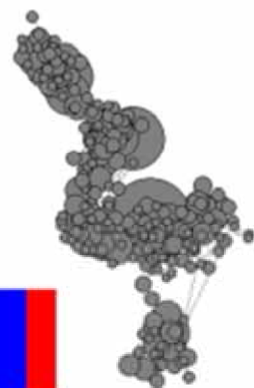
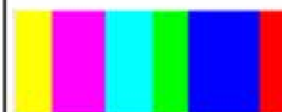


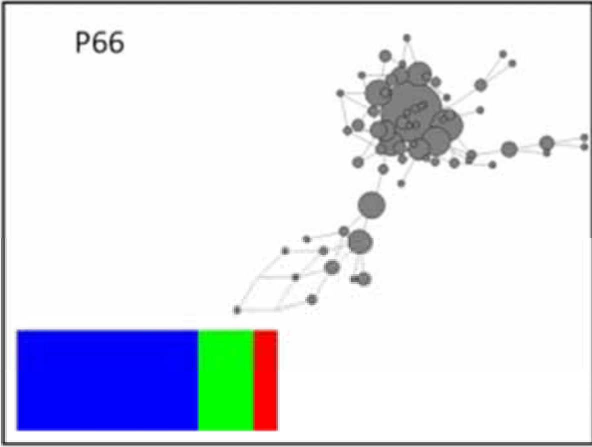









P48


P50


P51


P52


P61


P62



P66

Major versus minor: a detailed profile of physicochemical range in hepatitis C quasispecies sequences from ultra-deep pyrosequencing

Abstract. Hepatitis C virus is known as the main cause of chronic liver problems. Disease progression, pathogenicity, resistance to therapy and other biological features are related to the virus quasispecies composition. The hypervariable region 1 is the most heterogeneity region in viral genome and commonly used to quasispecies analysis. Sequences from baseline HVR1 genotype 3 were obtained using ultradeep pyrosequencing. Analysis of amino acid composition, physical and chemical properties, second structure composition and epitope analysis of this region, showed that despite highly sequence flexibility, there is consensus of epitope and coil/turns positions. This indicates that these regions may be essential to virus infection and points that HVR1 is an important marker of immune response in infection.

INTRODUCTION

Hepatitis C virus (HCV) presents a unique mechanism of persistence and host immune system evasion, which combines imunomodulation and viral high variability (7, 10, 18). HCV is classified in genotypes (numbers 1-7) and several subtypes (defined by letters)(12). Stand art therapy is changing from the non-specific Interferon based to the direct antiviral agents (DAA). However, the novel treatments are still expensive, present several side effects and are more effective to genotype 1 virus(4). Therefore, focus to resistance to therapy researches are changing from genotype 1 to 3, which still has not specific target drug developed and presents the higher rates of therapy failure to the new drugs (26).

The great number of different genome variants circulating in a single host, called quasispecies (15), is result of the combination of a high replication rate and the lack of proofreading activity of the viral RNA dependent-RNA polymerase (NS5B) (6). Quasispecies behavior in the host drives virus pathogenesis(22), disease progression(14), response to therapy(17) and viral evolution (2). Along HCV genome, there are regions of more or less conservation (27). Hypervariable regions (HVRs) are concentrated mainly in the Envelope 2 (E2) protein and they play an important role in evasion of immune response (11, 28). The HVR1 is part of the aminoterminal of E2 and contains epitopes for B-cell recognition (1, 25). Despite immune evasion, this region also participates in virus binding and entry (9). Therefore, HVR1 changes implicates in several steps of viral life cycle and despite the high flexibility of sequence, biochemical and structural properties are conserved(3).

This study analyzed the HVR1 phisico- and biochemical properties between major and minor quasispecies obtained from ultra-deep pyrossequencing in baseline samples. Molecular weight, hydrophobicity, isoeletric point, second structure changes and epitope mapping detailed how HVR1 changes between the variants in the same sample and between samples.

MATERIAL AND METHODS

Patients and samples. Baseline sera samples from 15 patients infected with Genotype 3, previously treated by combined therapy IFN+Ribavirin for 24 weeks were used. Response to therapy was determined by detection of viral RNA, classifying patients in three response groups: sustained virological responders (SVR), if HCV-RNA was detected only before treatment administration; non-responders (NR), if it was detected before, during and after end of treatment; and at last relapses (ETR), if viral RNA was not detected during, but became detectable in some time point after the end of therapy. Patients with co-infection with other virus which causes liver damage (ex. HBV) or immune deficiency (ex. HIV) were excluded.

All patients were enrolled at Sao Jose do Rio Preto city Blood Center and consented participation, according to Brazilian ethical rules in human researches.

HCV-HVR1 amplification and consensus sequencing. RNA from 100 µL sera was extracted by MagnaPureLC kit and MagnaPure Automated System (Roche, Inc., Indianapolis, IN). Eluted RNA was submitted to cDNA synthesis by cDNA SuperScript Vilo (Invitrogen, by Life Technologies). Amplification of HVR1 by PCR and Nested-PCR was performed using PerfeCta SYBR Green FastMix (Quanta Biosciences) LightCycler 480 instrument (Roche,Inc.) with primers and protocol previously described (21). Positive amplification samples were selected by melting curve analysis. Consensus sequences was obtained by BigDye v3.1 chemistry sequencing kit (Applied Biosystems, Foster City, CA) in automated sequencer (3130xl genetic analyzer; Applied Biosystems).

Ultra-deep pyrossequencing. First PCR products from HVR1 successful amplified samples were used for a new amplification round using primers from Nested-PCR modified with MID tags and adaptors. Fourteen set of primers were used and their sequence are described in Table S1 and S2. PCR products were purified by E-Gel® SizeSelect™ (Invitrogen, by Life Technologies). These purified products quality and quantity were verified in 2100 Bioanalyzer (Agilent DNA 1000 Kit, Agilent Technologies). Based on this quantification result, samples were diluted and pooled together in equimolar ratios. This mixture was submitted to pyrosequencig according to Roche/454 GS Junior (454 Sequencing, Roche) protocol for amplificon, described by manufacturer.

Data correction/filtering and analysis. Sequences extracted from 454 machine were corrected and filtered by pipeline developed, based on previous experiments using same protocol (23). Mega 5.2 software (24) was used for haplotypes alignment, sequences translation and singletons were removed for further analysis using Locqspec software (16). Mutation profile

was analysed using Highlighter tools at HIV databases from Los Alamos National Laboratory (http://www.hiv.lanl.gov/content/sequence/HIV/HIVTools.html). Dominant sequence was used as the master sequence. Statistical hypothesis tests were calculated using Minitab 16® (Quality. Analysis. Results.®, Minitab, Inc.) and p-values < 0.05 were considered significant. Molecular weight, isoeletric point (pI) and GRAVY (grand average of hydropathicity) were predicted using PROTPARAM(29). Second structure predictions were performed using Antheprot package(5) by Garnier methodology(8). For recognition of B-cell epitopes, it was used Bepipred tool(8).

RESULTS

From the fourteen samples, 8 were male and 7 female of 28-67 years old. Based on therapy outcome, 6 patients were responders, 6 were non-responders and 3 relapses. Pirosequencing pipeline processing resulted in 27,398 sequences. Total number of aminoacid haplotypes varied from 1 (P31 and P46) until 105 (P28), with no relation to therapy outcome verified by statistical test (ANOVA with Turkey and Fisher post hoc test).

Predicted molecular weight of the HVR1 region ranged from 10,67KDa (P28) to 11,13KDa (P47) (Figure 1A). Isoeletric point (pI) changed from 6 (P28) to 9.7 (P30) (Figure 1B) and GRAVY scale from 0.912 (P46) to 0.571 (P28) (Figure 1C). Those three parameters changed significantly between patients and between response to IFN groups, according to Kruskall-Wallis test with Dunn's post hoc test.

Highlighter plot of the mutations points were constructed for all samples (not shown). Alignment showed that amino acid changes were clustered in carboxiterminal of the sequences; also, changes for amino acids of different profiles of hydrophobicity or charge were rare in aminoterminal region. This pattern of substitution was not observed only in two patients samples (P61 and P66).

Hydrophobicity prediction along the HVR1 sequences showed that the in general this region presents high hydrophobic profile, mainly in central region (Figure 2). Inside the same patient, hydrophobicity range along the sequence was different. In samples from patients P30, P51, P61, 62 and P66 this range was distributed along the sequence, but in others the variation was greater in the carboxiterminal.

The second structure prediction indicates that despite the high sequence variability, most of the sequences presents similar conformation with predominant beta-sheet and conserved positions of coil/turns. Similar to the hydrophobicity behavior, sequences from patients P30, P51, P61, 62 and P66 showed a distribution of structure forms along the HVR1 region when compared to the master sequence.

Potential epitopes identified varied from one to three in carboxiterminal region, usually between position 63-75, 85-89 and 99-101. These positions seem to be high conserved in all the genotype 3 patients, although the number of epitopes changed between patients. Also, there was variation of epitopes number even inside the same patient.

DISCUSSION

HCV mutation rate is estimated to be 0.1 to 0.2 nucleotide mutations per genome site per year, therefore amino acid substitutions are common over a period of one year(13). The hypervariable 1 exhibit the highest degree of genetic heterogeneity in whole HCV genome (27) and represents a multitask region related to cell binding, entry(30) and immune system escape (1).

This last function is important to virus persistence mechanism(18), which is characteristic of HCV infection chronicity. To evaluate how host immune system could be working on the HVR1 region in patients with HCV genotype 3, our work analyzed several features of the

amino acid sequence in this region from detailed quasispecies profile from ultra-deep pyrosequencing.

First, overall molecular weight, pI and GRAVY (Figure 1) from the sequences indicated that HVR1 can be flexible when considered the size, and inside the same patient is possible to find variants that present more acid or more basic affinity. GRAVY analysis demonstrated that this region is naturally hydrophobic, which is consistent to the binding/interaction function.

Hydrophobic profile along the genome (Figure 2), demonstrated that in most patients there is a strong pressure along the aminoterminal of this region, which explains the conserved profile. However, it is important to notice that between patients samples, the general hydrophobic distribution along the sequence seems similar, which is related to the conservation of secondary structure between samples (19).

General secondary structure seems to be conserved in HVR1, even if the amino acid sequence shows high flexibility (Figure 3). Although small changes are observed all along the sequence, in general turns/coil distributions are conserved. This observation points to the importance of the folding of this region to maintain the E2 functional. Also, it is possible to discern a high conservation of three epitopes positions (Figure 4), even in sequences from different patients. This explains why there is cross-reaction of neutralizing antibodies recognition between HVR1 of different genotypes (20).

In overall, our study indicates that the major sequences and the minor quasispecies variants can have some flexibility in properties and the behavior is not similar in different patients. Also, although the immune recognition seems to concentrate in the carboxiterminal region, it does not mean that this region is more or less variable according to therapy to response type. For the last, HVR1 shows high sequence variability but features of hydrophobicity and secondary structure prediction seems to be conserved.

Acknowledgments

FAPESP and CAPES

FIGURES

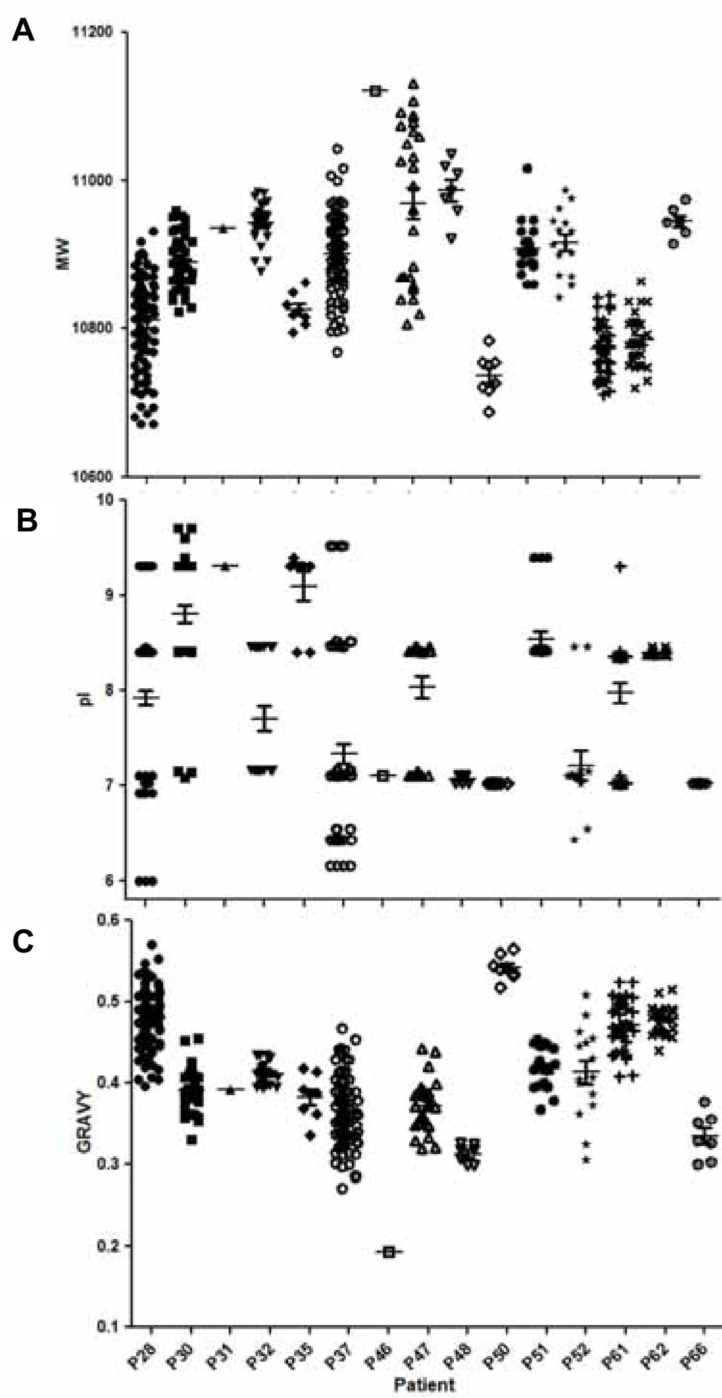
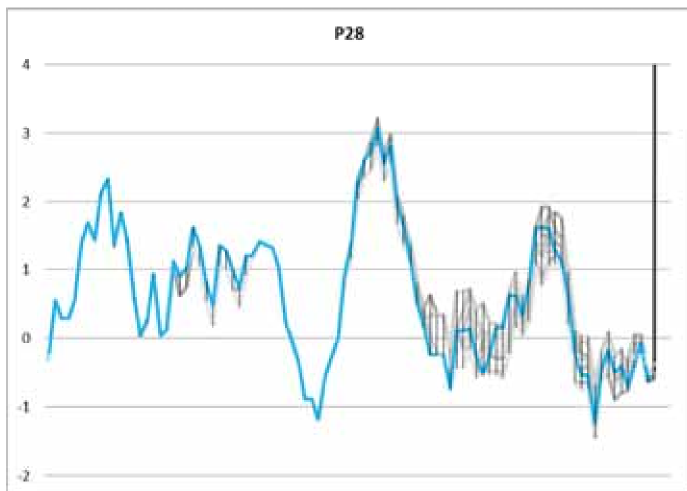
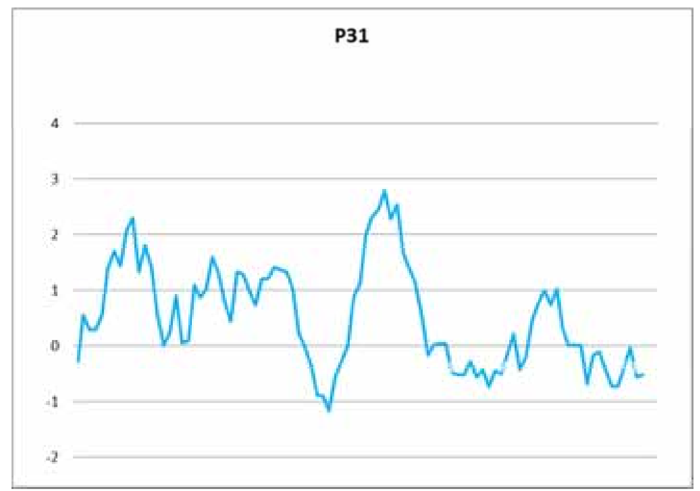
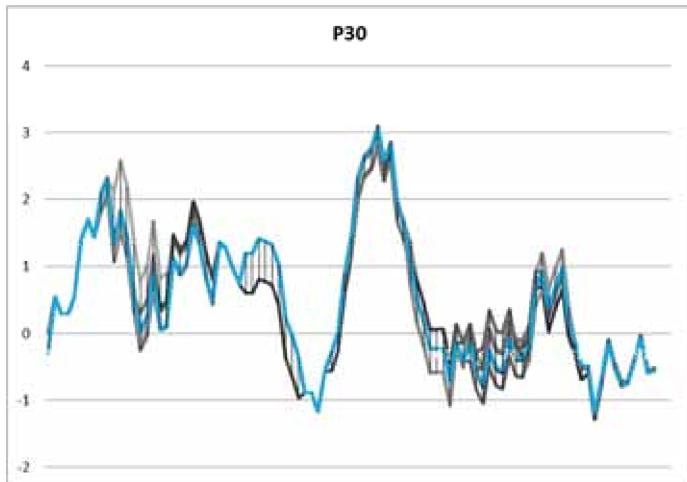
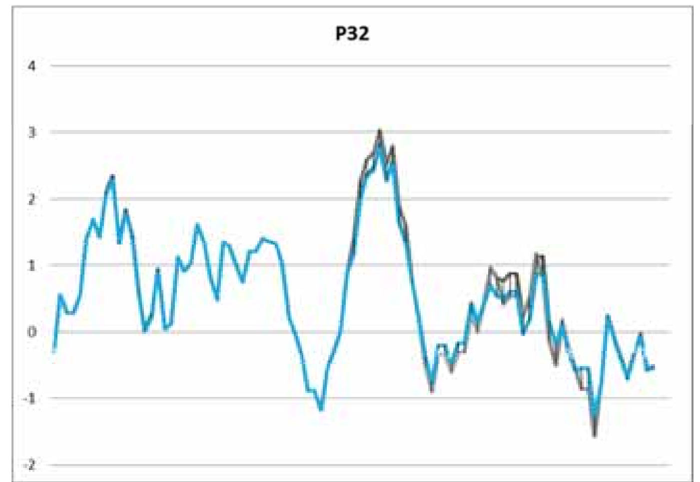


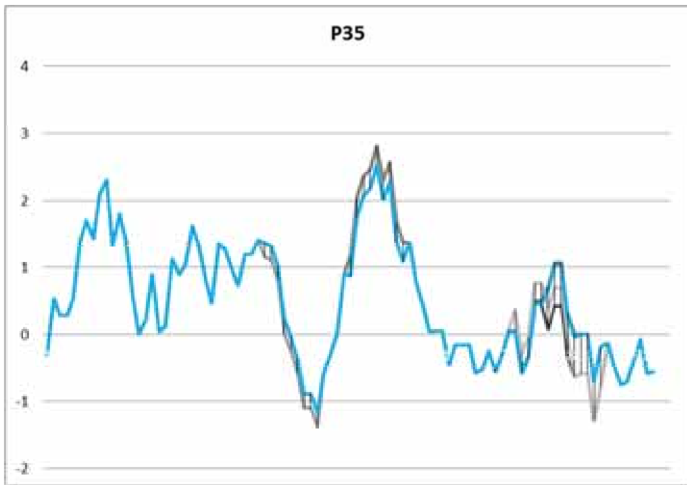
Figure 1. Results of ProtPram Plots. Each dot represent the value of total molecular weight - MW (A), Isoeletric point - pI (B) and Hydropacithy - GRAVY(C) calculated for each variant.

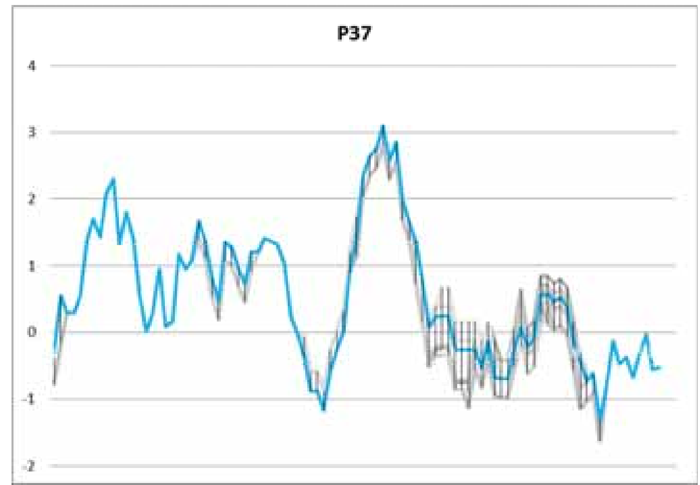


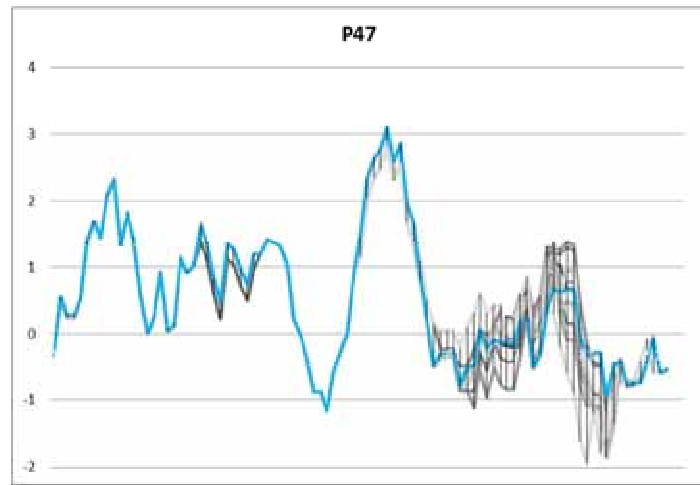




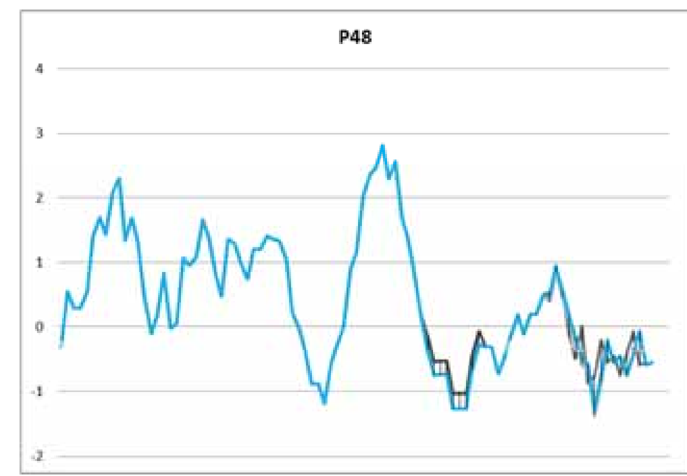


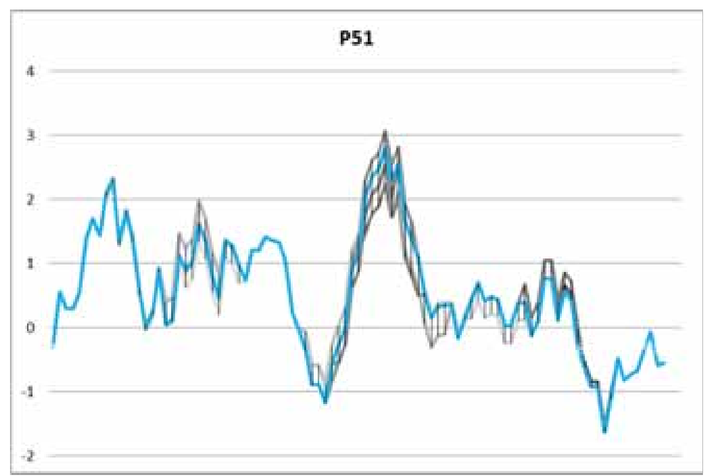


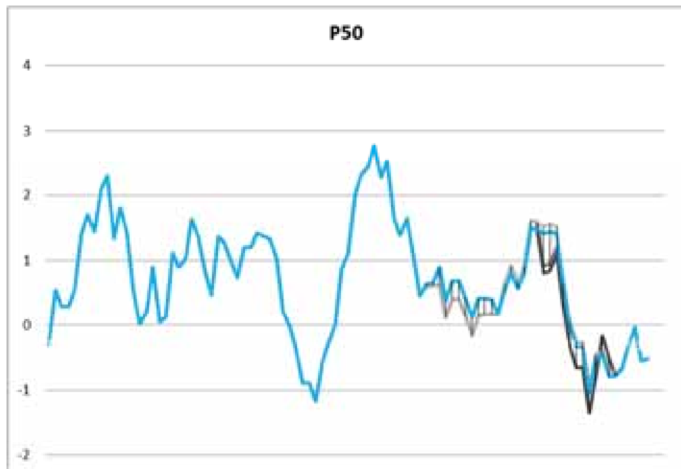


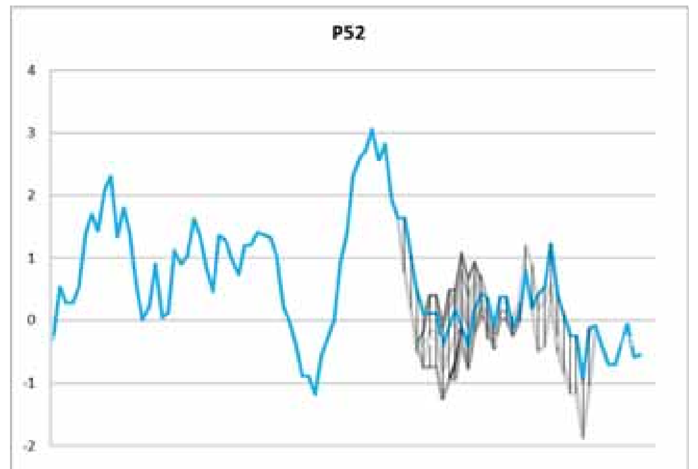


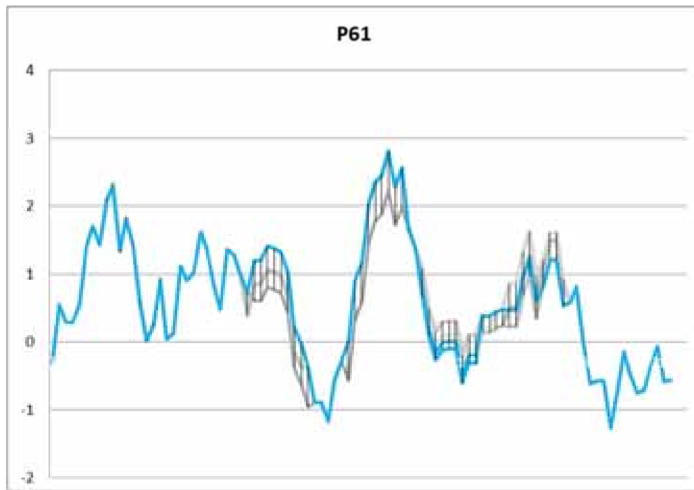
1 101 1 101

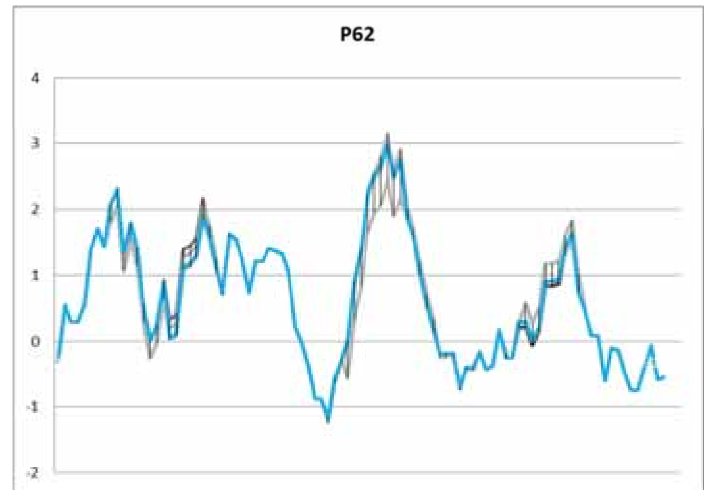






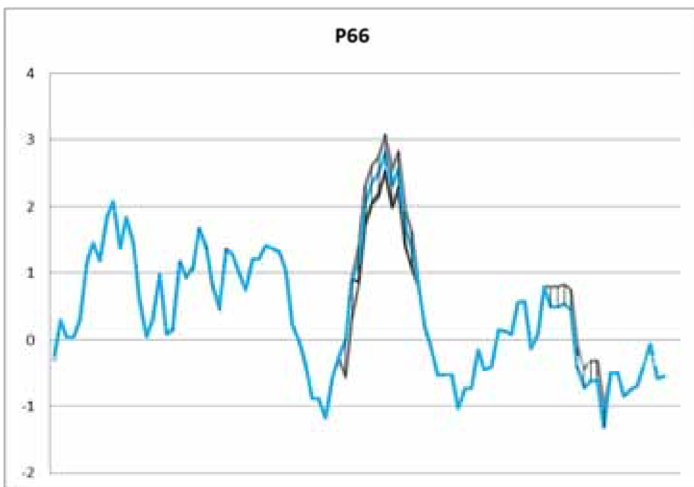






1

101



1

101

Figure 2. Hidrophobicity in each aminoacid position (x-axis) along the quasispecies HVR1 sequence, calculated by Kyte and Doolitle methodology (y-axis) . In blue: major variant, in grey: minor variants values.

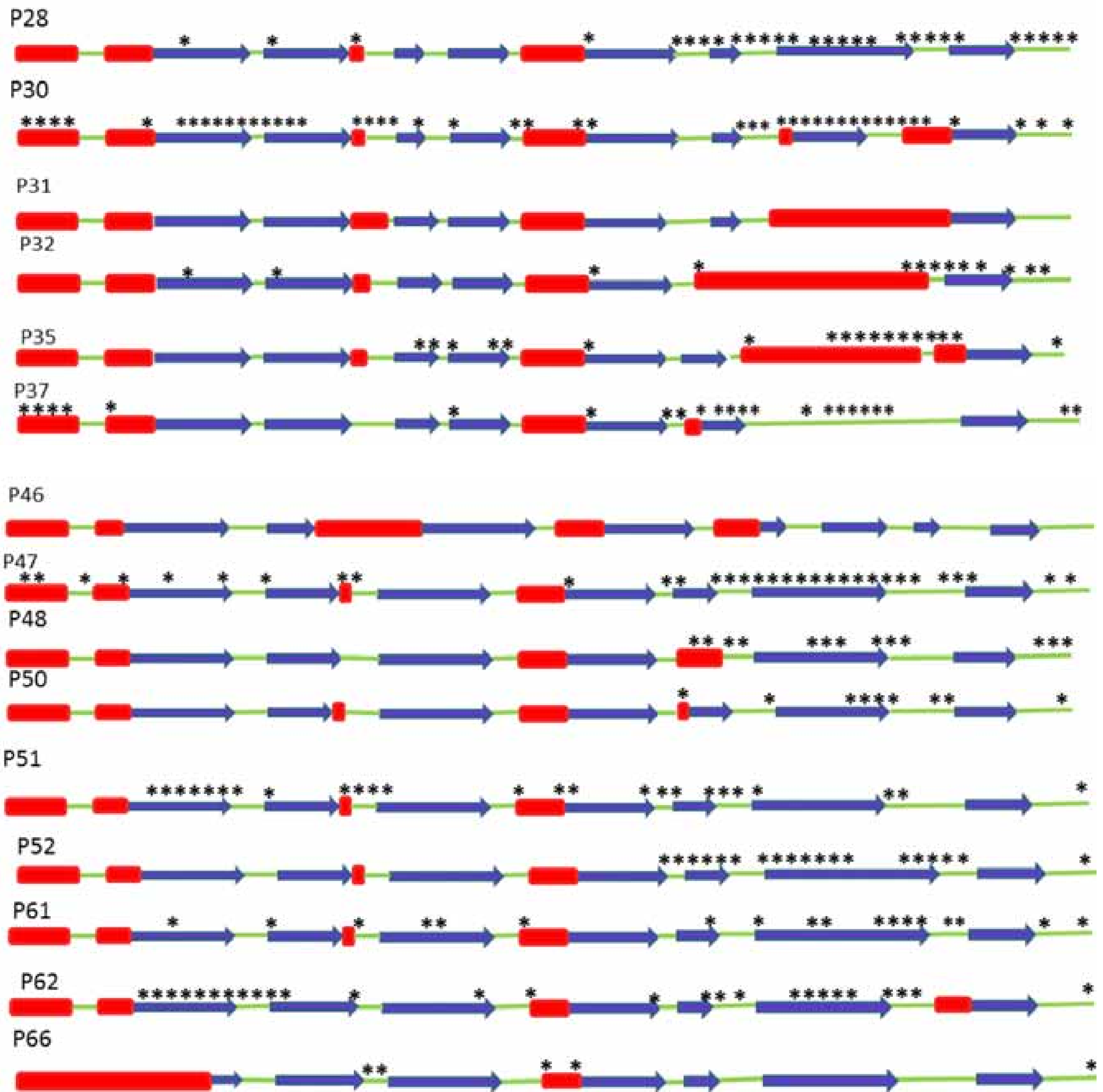


Figure 3. Secondary structure prediction of the major quasispecies (red square: helix; blue arrow: sheet and green line: turn/coil. * represents positions of change in minor variants

Figure 4. Representations of the major sequence epitopes positions and diversifications in minor quasispecies, identified by Bepipred. The major variant is represented on top with the epitopes in red. The following lines presents the epitopes found that did not match the major sequence epitope

REFERENCES

1. Bankwitz, D., E. Steinmann, J. Bitzegeio, S. Ciesek, M. Friesland, E. Herrmann, M. B. Zeisel, T. F. Baumert, Z. Y. Keck, S. K. Foung, E. I. Pecheur, and T. Pietschmann. 2010. Hepatitis C virus hypervariable region 1 modulates receptor interactions, conceals the CD81 binding site, and protects conserved neutralizing epitopes. J Virol 84:5751-5763.
2. Bull, R. A., F. Luciani, K. McElroy, S. Gaudieri, S. T. Pham, A. Chopra, B. Cameron, L. Maher, G. J. Dore, P. A. White, and A. R. Lloyd. 2011. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. PLoS Pathog 7:e1002243.
3. Callens, N., Y. Ciczora, B. Bartosch, N. Vu-Dac, F. L. Cosset, J. M. Pawlotsky, F. Penin, and J. Dubuisson. 2005. Basic residues in hypervariable region 1 of hepatitis C virus envelope glycoprotein e2 contribute to virus entry. J Virol 79:15331-15341.
4. Conteduca, V., D. Sansonno, S. Russi, F. Pavone, and F. Dammacco. 2013. Therapy of chronic hepatitis C virus infection in the era of direct-acting and host-targeting antiviral agents. J Infect.
5. Deleage, G., C. Combet, C. Blanchet, and C. Geourjon. 2001. ANTHEPROT: an integrated protein sequence analysis software with client/server capabilities. Computers in biology and medicine 31:259-267.
6. Domingo, E., V. Martin, C. Perales, A. Grande-Perez, J. Garcia-Arriaza, and A. Arias. 2006. Viruses as quasispecies: biological implications. Current topics in microbiology and immunology 299:51-82.
7. Gale, M., Jr., and E. M. Foy. 2005. Evasion of intracellular host defence by hepatitis C virus. Nature 436:939-945.
8. Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 120:97-120.
9. Guan, M., W. Wang, X. Liu, Y. Tong, Y. Liu, H. Ren, S. Zhu, J. Dubuisson, T. F. Baumert, Y. Zhu, H. Peng, L. Aurelian, P. Zhao, and Z. Qi. 2012. Three different functional microdomains in the hepatitis C virus hypervariable region 1 (HVR1) mediate entry and immune evasion. J Biol Chem 287:35631-35645.
10. Horner, S. M., and M. Gale, Jr. 2013. Regulation of hepatic innate immunity by hepatitis C virus. Nat Med 19:879-888.
11. Kato, N., Y. Ootsuyama, S. Ohkoshi, T. Nakazawa, H. Sekiya, M. Hijikata, and K. Shimotohno. 1992. Characterization of hypervariable regions in the putative envelope protein of hepatitis C virus. Biochem Biophys Res Commun 189:119-127.
12. Kuiken, C., and P. Simmonds. 2009. Nomenclature and numbering of the hepatitis C virus. Methods Mol Biol 510:33-53.
13. Kurosaki, M., N. Enomoto, F. Marumo, and C. Sato. 1993. Rapid sequence variation of the hypervariable region of hepatitis C virus during the course of chronic infection. Hepatology 18:1293-1299.
14. Lopez-Labrador, F. X., S. Ampurdanes, M. Gimenez-Barcons, M. Guilera, J. Costa, M. T. Jimenez de Anta, J. M. Sanchez-Tapias, J. Rodes, and J. C. Saiz. 1999. Relationship of the genomic complexity of hepatitis C virus with liver disease severity and response to interferon in patients with chronic HCV genotype 1b infection [correction of interferon]. Hepatology 29:897-903.
15. Martell, M., J. I. Esteban, J. Quer, J. Genesca, A. Weiner, R. Esteban, J. Guardia, and J. Gomez. 1992. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. J Virol 66:3225-3229.

16. Marucci, E. A., G. F. Zafalon, A. C. Jardim, L. H. Yamasaki, C. Bittar, P. Rahal, and J. M. Machado. 2008. Routine libraries for pattern recognition in quasispecies. Genet Mol Res 7:970-981.
17. Neumann, A. U., N. P. Lam, H. Dahari, D. R. Gretch, T. E. Wiley, T. J. Layden, and A. S. Perelson. 1998. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. Science 282:103-107.
18. Pavio, N., and M. M. Lai. 2003. The hepatitis C virus persistence: how to evade the immune system? J Biosci 28:287-304.
19. Penin, F., C. Combet, G. Germanidis, P. O. Frainais, G. Deleage, and J. M. Pawlotsky. 2001. Conservation of the conformation and positive charges of hepatitis C virus E2 envelope glycoprotein hypervariable region 1 points to a role in cell attachment. J Virol 75:5703-5710.
20. Prentoe, J., T. B. Jensen, P. Meuleman, S. B. Serre, T. K. Scheel, G. Leroux-Roels, J. M. Gottwein, and J. Bukh. 2011. Hypervariable region 1 differentially impacts viability of hepatitis C virus strains of genotypes 1 to 6 and impairs virus neutralization. J Virol 85:2224-2234.
21. Ramachandran, S., D. S. Campo, Z. E. Dimitrova, G. L. Xia, M. A. Purdy, and Y. E. Khudyakov. 2011. Temporal Variations in the Hepatitis C Virus Intra-Host Population during Chronic Infection. J Virol.
22. Rau, M., K. Baur, and A. Geier. 2012. Host genetic variants in the pathogenesis of hepatitis C. Viruses 4:3281-3302.
23. Skums, P., Z. Dimitrova, D. S. Campo, G. Vaughan, L. Rossi, J. C. Forbi, J. Yokosawa, A. Zelikovsky, and Y. Khudyakov. 2012. Efficient error correction for next-generation sequencing of viral amplicons. BMC Bioinformatics 13 Suppl 10:S6.
24. Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution 28:2731-2739.

25. Taniguchi, S., H. Okamoto, M. Sakamoto, M. Kojima, F. Tsuda, T. Tanaka, E. Munekata, E. E. Muchmore, D. A. Peterson, and S. Mishiro. 1993. A structurally flexible and antigenically variable N-terminal domain of the hepatitis C virus E2/NS1 protein: implication for an escape from antibody. Virology 195:297-301.
26. Tapper, E. B., and N. H. Afdhal. 2013. Is 3 the new 1: perspectives on virology, natural history and treatment for hepatitis C genotype 3. J Viral Hepat 20:669-677.
27. Weiner, A. J., M. J. Brauer, J. Rosenblatt, K. H. Richman, J. Tung, K. Crawford, F. Bonino, G. Saracco, Q. L. Choo, M. Houghton, and et al. 1991. Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. Virology 180:842-848.
28. Weiner, A. J., H. M. Geysen, C. Christopherson, J. E. Hall, T. J. Mason, G. Saracco, F. Bonino, K. Crawford, C. D. Marion, K. A. Crawford, and et al. 1992. Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. Proc Natl Acad Sci U S A 89:3468-3472.
29. Wilkins, M. R., E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel, and D. F. Hochstrasser. 1999. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 112:531-552.
30. Zhou, Y. H., Y. K. Shimizu, and M. Esumi. 2000. Monoclonal antibodies to the hypervariable region 1 of hepatitis C virus capture virus and inhibit virus adsorption to susceptible cells in vitro. Virology 269:276-283.

3. Conclusões Gerais

- As quasispecies da HVR1 em amostras pré-tratamento do genótipo 3 podem se apresentar de forma conservada (homogênea) ou altamente variável quanto a sequência de nucleotídeos e aminoácidos;
- A diversidade da HVR1 não apresentou correlação com a resposta ao tratamento ao IFN;
- A região carboxiterminal se apresentou como a microrregião de maior variabilidade na maioria dos pacientes;
- Estrutura secundária e epítopos são conservados quanto à posição, mas não quanto ao número e tamanho na HVR1;
- A análise das quasispecies da HVR1 é importante para entender a reação do vírus a pressão do sistema imune do hospedeiro.