Published in final edited form as:

J Virol Methods. 2013 August ; 191(2): 105–112. doi:10.1016/j.jviromet.2013.04.010.

Comprehensive Cloning of Patient-derived 9022-bp Amplicons of Hepatitis C Virus

Yang Lu^{a,c}, Yanjuan Xu^a, Adrian M. Di Bisceglie^{a,b}, and Xiaofeng Fan^{a,b,*}

^aDivision of Gastroenterology and Hepatology, Department of Internal Medicine, Saint Louis University School of Medicine, St. Louis, Missouri 63104, U.S.A

^bSaint Louis University Liver Center, Saint Louis University School of Medicine, St. Louis, Missouri 63104, U.S.A

^cDepartment of Internal Medicine, Yangzhong People Hospital, 235 Yangzi Road, Yangzhong 212200, Jangsu Province, China

Abstract

The instability of recombinant clones accommodating large or full-length viral genomes is frequently a technical challenge in RNA virus research. In an attempt to establish a rapid plasmid-based reverse genetics system that utilizes long RT-PCR technique (LRP), similar difficulty was encountered in the cloning of 9022-bp LRP amplicon. All HCV genotype 1a strains used for LRP cloning showed a remarkable difference in terms of cloning stability. Subsequent analysis revealed the predictive value of phylogenetic positions in determining the cloning stability. Putative *E. coli* promoters on the HCV genome might be responsible for such cloning difference. An exhaustive exploration, testing nearly one hundred cloning protocols, did not reveal a general approach that can achieve stable cloning for all HCV 1a strains. The selection of appropriate strains, guided by phylogenetic analysis, appears to be necessary prior to the construction of infectious HCV 1a clones. These observations are not only valuable for potentially establishing an HCV 1a cell culture model but also have general implications for other RNA viruses due to concern about cloning instability.

Keywords

hepatitis C virus; long RT-PCR; cloning

1. Introduction

Hepatitis C virus (HCV) is responsible for most post-transfusional and community-acquired hepatitis (Alter et al., 1999). HCV is an enveloped single-stranded, positive-sense RNA virus classified into the *Flaviviridae* family (Choo et al., 1989). The viral genome comprises an approximate 9600 bp RNA molecule that is divided into three regions: 5' untranslated region (UTR), a single large open reading frame and a 3'UTR. Despite considerable research over the past three decades, the molecular events under HCV infection are not

*Corresponding authors: Division of Gastroenterology and Hepatology, Department of Internal Medicine, Saint Louis University School of Medicine, 3635 Vista Avenue, St. Louis, MO 63110, Phone: +1 314-977-7833, Fax: +1 314-577-8125, fanx@slu.edu (X. Fan).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

completely understood. A major obstacle for HCV research remains the lack of reliable *in vitro* models supporting the full life cycle of HCV (Tellinghuisen et al., 2011). Only one particular HCV genotype 2a strain, JFH1, is able to be propagated in Huh-7-based cells (Wakita et al., 2005; Zhong et al., 2005). The reason why only HCV JFH1 strain works in cell culture remains elusive, however, it seems apparent that the HCV genome itself must play a role. To search for other culturable HCV genomes, especially from HCV 1a, the most prevalent HCV genotype in the world, a rapid plasmid-based reverse genetics system might be valuable by utilizing long RT-PCR technique that reproducibly amplifies nearly full-length HCV genomes from clinical samples (Fan et al., 2006; Zhou et al., 2007; Xu et al., 2008). In this approach, the cloning of LRP product is a central process. However, unexpected difficulty has been encountered in the cloning of LRP product from some HCV strains, presumably due to the instability of recombinant clones. Although this phenomenon has long been recognized in the construction of infectious RNA virus clones (Ruggli, et al., 1999), there are few studies that have investigated systematically its intrinsic mechanisms and alternative cloning approaches from a technical point of view. The current study conducted exhaustive experimentation to evaluate nearly all available options to enhance the cloning stability. A general guideline was also suggested for the selection of viral strains in the setting of rapid reverse genetics.

2. Materials and Methods

2.1. Serum samples

A total of five serum samples were used in the current study. Samples #1106, #1701, #1709 and #4701, all HCV genotype 1a, were chosen from a previous study based on their intra-host viral diversity and sample volumes available for experimental optimization (Fan et al., 2009). The serum sample H77, representing the prototype of HCV genotype 1a, was a gift from Dr. Robert Purcell (National Institutes of Health, U.S.A.). All samples were aliquoted and stored at -80°C until use.

2.2. LRP and basic cloning protocol

The amplification of nearly full-length HCV genome from serum samples utilized an experimental protocol detailed previously (Fan et al., 2006). In brief, an aliquot of $9.4\text{ }\mu\text{l}$ of extracted serum RNA was mixed with $10.6\text{ }\mu\text{l}$ of reverse transcription (RT) reaction matrix consisting of 1x SuperScript III buffer, 10 mM DTT, $1\text{ }\mu\text{M}$ QR2 (reverse primer), 2 mM dNTPs (Life Technologies, Grand Island, NY), 20 U of Rnasein Ribonuclease Inhibitor (Promega, Madison, WI), 200 U of SuperScript III (Life technologies) and 5 U of AMV (Promega). After incubation at 50°C for 75 min, the reaction was inactivated by heating at 70°C for 15 min. An aliquot of $5\text{ }\mu\text{l}$ RT reaction was applied for the first round of PCR that contained 1.25 mM $\text{Mg}(\text{OAc})_2$, 1x XL PCR buffer, 2 mM of dNTPs (Life technologies), $0.4\text{ }\mu\text{M}$ of Trnc-21, $0.4\text{ }\mu\text{M}$ of each primer (WF33 and QR2) and 2 U of rTth XL DNA polymerase (Life technologies). Cycle parameters were programmed as 94°C for 1 min followed by the first 10 cycles of 94°C for 30 sec, 65°C for 30 sec and 68°C for 9 min and the final 20 cycles in which the annealing temperature was reduced to 60°C . The reaction was ended with 10-min incubation at 72°C . An aliquot of $2\text{ }\mu\text{l}$ of the first round of PCR product was used for the second round of PCR amplification with primers WF5 and WR55. Cycle parameters were the same as the first round PCR except the annealing/elongation temperature was changed to 72°C for the first 10 cycles and 68°C for the last 20 cycles, respectively.

After the spin purification using QIAquick PCR Purification Kit (Qiagen, Valencia, CA), the LRP product was digested with PacI/FseI (New England BioLabs, Ipswich, MA), followed by gel purification with QIAEX II Gel Extraction Kit (Qiagen). An aliquot of 226 ng of LRP

product was subjected to ligation containing 25 ng of PacI/FseI digested pClone vector (1:3 vector/insert ratio) (Fan et al., 2006), 1 x ligation buffer and 1 μ l of T4 DNA ligase (5U/ μ l) (Life Technologies) in a total of 10 μ l reaction volume. After the incubation at 14°C overnight, 1.5 μ l of 1:5 diluted ligation product was used directly for the electroporation into 25 μ l of ElectroMAX™ DH10B™ T1 Phage-Resistant electrocompetent *E. coli* cells (Life Technologies), followed by incubation at 37°C for 1 hour. About 50 μ l of cell suspension was casted on lysogeny broth (LB) plates containing 30 μ g/mL kanamycin. The plate was incubated at 37°C overnight and thirty recombinant clones were picked at random for the screening of the insert by colony PCR or enzyme digestion (Fan et al., 2006). It should be noted that all parameters related to the basic cloning protocol were fixed to facilitate the comparison of various modifications in cloning protocols.

2.3. Modifications of cloning procedure

The following sections describe every protocol's modification conducted to enhance cloning stability. These methods were tested individually or in combination, so that nearly 100 experimental protocols were estimated in the current study.

2.3.1. Physical or chemical modification during the culture—Two temperatures, 24°C (room temperature) and 30°C were applied in two steps during cloning, the 1-hr recovery shaking after electroporation and overnight LB plate incubation. In addition, the electroporation product was casted on LB plates containing various concentrations of glucose (Sigma, St. Louis, MO) from 0.36% to 1%, as the addition of glucose may repress basal expression of potentially toxic genes from either the vector or the insert (Zhang et al., 2003; Studier et al., 2005).

2.3.2. Switch of insert direction—It is commonly hypothesized that cloning instability results from the cryptic transcription of toxic domains from the insert. One approach to reduce such undesirable transcription is to switch cloning direction of the insert. In the current study, this was achieved by simply exchanging the restriction sites from the primers WF5 and WR55, which were used for the second round of LRP amplification.

2.3.3. LRP purification by gelase—To compare the quality of the templates used for ligation and subsequent cloning, the LRP product was purified using an alternative method. In brief, the LRP amplicon was gel purified by using gelase (Epicentre, Madison, WI) according to the manufacturer's instruction. This purification procedure was used to replace the two purification steps before (QIAquick PCR Purification) and after (QIAEX II Gel Extraction) the restriction digestion of LRP amplicon.

2.3.4. Creation of minimal pClone vectors—To eliminate possible toxic transcription from the vector, two minimal pClone vectors were created. These vectors retained only elements essential for plasmid propagation, i.e., the origin of replication (OR), an antibiotic gene and a multiple cloning site (MCS). In doing so, the OR domain was digested with PacI/BspHI (New England Biolabs) from the pClone vector (3030 bp), a pBR322-derived vector (Fan et al., 2006). The kanamycin gene was amplified from the pClone with primers KF1 (5'-ata gct ggg ttc atg ata gtt tgc gca acg ttg ttg -3') and KR (5'-gga tct gac gtt aat taa gcc gat ttc ggc cta ttg gt -3'), containing restriction sites BspHI and PacI, respectively. After the digestion and gel purification, the PCR product was ligated with the OR fragment, followed by transformation into Library Efficiency DH5 α Competent Cells (Life Technologies). Correct clones were confirmed by full sequencing. The plasmid was named pCloneK, 2222 bp in length. Similarly, pCloneA (1946 bp) was created by the replacement of kanamycin with the ampicillin gene, which was amplified from the pUC19 vector with primers AF (5'-

ata gct ggg ttc atg att caa ata tgt atc cgc aca tg -3') and AR (5'-gga tct gac gtt aat taa tta cca atg ctt aat cag tga g -3'). Both vectors were then used for LRP cloning as described above.

2.3.5. Estimation of other *E. coli* cells—Various *E. coli* cells were tested for their ability to accommodate plasmids with large inserts. Among seven *E. coli* cells tested, four were from Competent Cell Selection Kit (Life Technologies), i.e., OmniMAX™ 2-T1R Cells, Mach1-T1R Cells, Stbl3™ Cells and TOP10 Chemically Competent *E. coli* Cells. ElectroMAX Stbl4 Electrocomp Cells (Life Technologies) was also tried. The other two were SURE competent cells (Agilent Technologies, Santa Clara CA) and CopyCutter EPI400 Electrocompetent cells (Epicentre). The pCloneA vector was used for the estimation of the SURE cells due to its intrinsic resistance to kanamycin.

2.3.6. Engineering transcription terminators into the vectors—To minimize the transcription of potential toxic genes, *E. coli* transcription terminators rrnB T1, T2 and tHP were engineered into both sides of the MCS of the pCloneK vector (Nohno et al., 1986, 1988; Orosz et al., 1991; Brown et al., 1993). In brief, the AsisI/BamHI fragments containing terminator sequences were assembled using PCR, followed by digestion and gel purification. The fragments were then used to replace the corresponding domain of pCloneK by standard molecular cloning, which generated two modified vectors, pCloneKR (2,565 bp) and pCloneKT (2,419 bp) (Figure 1). The pCloneKR contained *E. coli* rrnB T1 and T2 on both sides of the MCS while the *E. coli* rrnB T1 and T2 on the left side of the MCS were replaced with tHP in the pCloneKT (Figure 1). All construction was confirmed by full sequencing. The restriction sites PacI and XbaI from new vectors were utilized for the cloning of LRP product, which was amplified using primers containing corresponding restriction sites.

2.3.7. The insertion of par element into the vectors—The partition (par) domain of plasmid pSC101 has been reported to stabilize plasmid propagation (Manen et al., 1990). This element (371 bp) was thus inserted into plasmids pCloneKR and pCloneKT by utilizing a single restriction site, BspHI, on the vectors. Briefly, the par domain was amplified from the plasmid pSC101 (ATCC, Manassas, VA) with primers parF (5'-ata gct ggg ttc atg aga cag taa gac ggg taa gcc -3') and parR (5'-gga tct gac gtc atg act cgg gca aat cgc tga ata -3'). The PCR product was digested with BspHI and then inserted into the BspHI site of pCloneKB or pCloneKT by standard cloning protocol. Correct recombinant plasmids were confirmed by enzyme digestion/sequencing and named pCloneB_par and pCloneKT_par.

2.3.8. BAC vectors—BAC vectors are well known for their capability to accommodate large inserts (She et al., 2003). In the current study, two commercial BAC vectors were tested for the cloning of LRP product. The first BAC vector was CopyControl pCC1BAC (BamH I Cloning-Ready) (Epicentre). The single BamHI site on the pCC1BAC was used to introduce the MCS domain amplified from the pClone vector with the primers MCSFBamHI (5'-ata gct ggg tgg atc ctt aat taa cga gcg atc gcg -3') and MCSRBamHI (5'-gga tct gac ggg atc ccc tgc agg ggc cga gat -3'). The single restriction site PacI was used for LRP cloning due to the existence of both FseI and XbaI sites on the pCC1BAC backbone. The LRP product was therefore amplified with primers containing only PacI sequence at the 5' ends. It should be noted that the ligation product was electroporated into TransforMax EPI300 electrocompetent *E. coli* cells (Epicentre), which was engineered specially for an auto-induction of plasmid copy numbers (Wild et al., 2002). The second commercial vector was the pEZ BAC (Lucigen, Middleton, WI). This vector is similar to pCC1BAC except for the existence of transcription terminators on both sides of the cloning site. The ligation product was electroporated into 10G SUPREME electrocompetent cells, a derivative of DH10B cells.

2.3.9. Ligation independent cloning (LIC)—The LIC creates long “sticky ends” by the treatment of T4 DNA polymerase in the presence of only a single nucleotide (Aslanidis et al., 1990). In the current study, the primers for the second LRP amplification were replaced with WF5T4P (5′-ctc acc tta att aac ata gtg gtc tgc gga acc ggt -3′) and WR55T4P (5′-ata cct ccc ttc tac aat ggc agc cct gcc tcc tct gg -3′). An aliquot of 2.28 μg (~0.5 pmol) purified LRP product mixed with 5 U of T4 DNA polymerase (New England Biolabs) in the presence of 2.5 mM dCTP (Invitrogen), incubated at room temperature for 30 minutes followed by heat inactivation for 20 minutes, which generated 18-bp sticky ends on the LRP product. Similarly, pCloneK was amplified fully with primers OriFT4P (5′-tat gtt aat taa ggt gag cgg tgt gaa ata ccg cac ag -3) and OriRT4P (5′-att gta gaa ggg agg tat ccg att tgc gcc tat tgg tt -3′). The resulting PCR product was treated with T4 DNA polymerase in the presence of dGTP, which created 18-bp sticky ends matching to the one from the LRP amplicon. The T4 DNA polymerase-treated amplicons from both the pCloneK vector and LRP were then annealed by the incubation at room temperature for 1 hour, followed by inactivation with the addition of 1 μl of 25 mM EDTA (pH 8.0). The annealing product was electroporated into ElectroMAX™ DH10B™ T1 Phage-Resistant electrocompetent *E. coli* cells (Life Technologies).

2.3.10. The incorporation of 7-Deaza-2′-deoxy-guanosine-5′-triphosphate (7-deaza-dGTP) during LRP amplification—7-deaza-dGTP is an analogue of dGTP used commonly in amplification and sequencing of DNA with a high content of GC, such as promoter regions (Jung et al., 2002; Frey et al., 2008). It is thus possible that the incorporation of 7-deaza-dGTP may impair potential activity of cryptic promoters. To test this hypothesis, the second round LRP was performed with a range of dGTP/7-deaza-dGTP (Roche Applied Science, Indianapolis, IN) ratios from 0% to 80%. Since the incorporation of 7-deaza-dGTP may result in invisibility of PCR product in ethidium bromide- stained agarose gel (Weiss et al., 1994), the LRP product was tested with alternative DNA stains, including Midori Green DNA Stain (a gift from Nippon Genetics Europe GmbH, Germany), 4′,6-diamidine- 2-phenylindole (DAPI) nucleic acid stain (Life Technologies), BlueView DNA stain (Sigma) and SYBR safe DNA stain (Life Technologies). The 7-deaza-dGTP-incorporated LRP was then subjected to basic cloning procedure.

2.3.11. Partial and whole plasmid mutagenesis—The pCloveK vector was used for partial or whole plasmid mutagenesis to create potential new vector mutants to accommodate large LRP insert. Partial plasmid mutagenesis focused on the domain of the origin of replication (OR) using error prone PCR with two different approaches. The first approach utilized Taq DNA polymerase (Life Technologies), the incorporation of dITP (New England Biolabs), the unbalanced dNTPs and the inclusion of manganese chloride (MnCl₂) in PCR reaction buffer. These modifications may generate a mutant library after the template amplification (Cirino et al., 2003). The second method used GeneMorph II Random Mutagenesis Kit (Agilent Technologies) in which the mutazyme II DNA polymerase has a high mutation rate with minimal mutational bias at each nucleotide sites along with the template. In either approach, the pCloneK was digested with PacI/BspHI to generate two fragments. The large fragment (1,220 bp), containing kanamycin gene, was gel purified and named fragment K. The OR domain, named fragment R, was amplified by error prone PCRs with primers OriFXbaI (5′-ata gct ggg ttc tag acg gtg tga aat acc gca cag -3′) and OriRBspHI (5′-gga tct gac gtc atg acc aaa atc cct taa cgt g -3′), followed by XbaI/BspHI digestion and purification. Three fragments, K (1200 bp), R (900 bp) and PacI/XbaI digested LRP, were ligated together followed by basic cloning procedure.

The whole plasmid mutagenesis was conducted with GeneMorph II Random Mutagenesis Kit only. The entire pCloneK vector was amplified with primers OriFXbaI and OriRPacI (5′-gga tct gac gtt aat taa gcc gat ttc ggc cta ttg gt -3′), followed by the procedure as

described above. In either partial or whole plasmid mutagenesis, it should be noted that 200 recombinant clones instead of 30 were screened for the existence of the LRP insert by colony PCR.

2.4. 454 sequencing

The LRP product of all five samples (#1701, #1709, #1106, #4701 and #H77) was subjected to pyrosequencing (454 sequencing). Briefly, 5 µg of gel-purified LRP product was used for the construction of fragment library with Rapid Library Preparation Kit (Roche Applied Science), followed by pyrosequencing on GS/FLX Titanium. Resulting sequencing reads were assembled to generate consensus sequences by reference mapping to H77 sequence (GenBank accession number: AF009606) in 454 software (Newbler).

2.5. Phylogenetic analysis

Phylogenetic analysis was used to explore the potential association between cloning stability and sequence clustering. Two phylogenetic trees were constructed with either partial HCV envelope sequences or nearly full-length HCV genomes from nine HCV strains done with LRP cloning in the current or previous studies (Zhou et al., 2007; Xu et al., 2008). Both trees were computed in using the MEGA 5 program with a Neighbor-Joining approach under the maximum composite likelihood model and an empirical value of among-site variation ($\alpha = 0.5$) (Tamura et al., 2011). The same model was also used for the calculation of genetic distances between each nearly full-length HCV 1a genome investigated in the current study and HCV genotype 1a prototype H77 strain.

2.6. Promoter analysis

Putative *E. coli* promoters (ECP) on HCV genomes were predicted using a neural network promoter program (Reese et al., 2001). The number of ECP with a score large than 0.9 was counted for all eight HCV LRP consensus sequences, including five sequences in the current study and three from previous studies, i.e., HCV strains Liv23 (DQ838739), Liv02 and Liv03 (Zhou et al., 2007; Xu et al., 2008). The LRP consensus sequences of Liv02 and Liv03 were derived respectively from three cloned sequences, EU677247, EU677248 and EU677249 for Liv02 and EU677253, EU677254 and EU677255 for Liv03. The calculation of ECP numbers was also made for 395 non-redundant HCV genotype 1a full coding sequences retrieved from HCV database (Kuiken et al., 2005). Potential clustering among 395 full coding sequences was explored by phylogenetic tree construction as described above. Statistical significance of ECP difference among phylogenetic clusterings was determined by two-tailed Student t test.

2.7. GenBank accession numbers

A total of nine LRP sequences, five consensus LRP sequences generated by 454 sequencing and four LRP sequences from the whole-plasmid mutagenesis, were deposited in the GenBank under the accession numbers JX472005 through JX472013.

3. Results

3.1. Difference of LRP cloning stability among HCV genotype 1a strains

LRP was performed successfully with all five HCV genotype 1a samples #1106, #1701, #1709, #4701 and H77. All LRP products were pyrosequenced at considerable depth. A single large contig was generated for each LRP product by the reference mapping. Together with the high read mapping rates, ranging from 96.2% to 98%, the combination of LRP and 454 sequencing represents a rapid and feasible approach for the determination of viral consensus sequence. In subsequent LRP cloning into pClone vector, however, only samples

#1701 and #1709 showed high positive rates of recombinant clones by colony PCR and/or enzyme digestion. Among thirty recombinant clones picked at random, less than three clones were positive for sample #1106 and none was confirmed for the insert in samples #4701 and H77 (Table 1). Cloning stability appeared not to be associated with viral loads or viral quasispecies diversity (Table 1). Instead, these strains showed differential genetic distances against H77 strain (Table 1). By including HCV genotype 1a strains with recorded positive rates in previous studies (Table 1), phylogenetic analysis showed all difficult-to-clone HCV strains to be clustered into clade 1 while highly cloneable strains were placed at either clade 2 or clade 3 (Figure 2). Exactly the same clustering was observed in the tree constructed with nearly full-length HCV genomes generated by 454 sequencing (data not shown).

3.2. Roles of experimental modifications on LRP cloning stability

Two difficult-to-clone samples, H77 and #4701, were used to test other cloning approaches or modifications. These methods were estimated individually or in combination thus generating nearly 100 experimental protocols. However, none of them gave significant improvement with regard to cloning stability (Table 2).

The plasmids pCloneK and pCloneA, constructed by removing unessential domains, showed comparable cloning stability with the pClone vector. However, both plasmids cannot accommodate the LRP insert from samples H77 and #4701.

A further modification generated the pCloneKR and pCloneKT vectors by engineering transcription terminators onto both sides of the cloning site (Figure 1). Again, these plasmids were as efficient as the parent pCloneK vector to accommodate the LRP insert from HCV samples in clades 2 and 3 but not clade 1.

Two BAC vectors, pCC1 and pEZ, slightly improved the cloning stability. Only one to three recombinant clones were positive in repeated experiment with either BAC vector (Table 2).

The incorporation of 7-deaza-dGTP did not affect the LRP amplification. However, when the ratio of 7-deaza-dGTP/dGTP was larger than 70%, the LRP product became invisible in ethidium bromide-stained agarose gel. Among other DNA stains, only BlueView resulted in visible staining in agarose gel to guide the excision of DNA bands. However, similar efficiency to ethidium bromide to stain regular LRP product was observed for these stains, including Midori Green DNA Stain, DAPI nucleic acid stain and SYBR safe DNA stain. The incorporation of 7-deaza-dGTP did not improve cloning stability for samples H77 and #4701 (Table 2). Such an incorporation even reduced positive rates of LRP cloning from samples in HCV 1a clades 2 or 3 (data not shown), indicating a toxic role of 7-deaza-dGTP in molecular cloning. It should be noted this observation excluded the possibility of the resistance to restriction enzymes by the incorporation of 7-deaza-dGTP (Grime et al., 1991), which was tested first by the digestion of 7-deaza-dGTP-incorporated PCR product containing restriction sites, including PacI, XbaI and AsisI (data not shown).

Finally, partial mutagenesis of pCloneK by error-prone PCR did not produce any vector mutants capable of accommodating LRP inserts. In contrast, 4 out of 200 screened recombinant clones were positive in the approach of whole plasmid mutagenesis. However, these four vectors were unable to maintain the cloning stability in LRP cloning of sample H77. All LRP inserts from these vectors were sequenced by gene-walking method (Table 3). Genetic analysis revealed 29 to 51 nucleotide substitutions in comparison with the H77 consensus sequence from the 454 sequencing. These substitutions were distributed randomly along the entire LRP amplicon without consistent clustering in a given domain.

3.3. Variation of putative ECP numbers among HCV genotype 1a strains

Putative *E. coli* promoter (ECP) was counted for each consensus sequence of HCV samples in the current study. The easy-to-clone samples (#1709 and Liv02) had fewer ECPs, however, high ECP number was not always associated with the cloning difficulty (Table 1). The samples #1701 and Liv03 had comparable ECP numbers with difficult-to-clone samples, i.e., H77, #1106 and #4701 (Table 1).

The phylogenetic tree with 395 full-length HCV genotype 1a sequences showed a similar topology with the tree based on partial HCV envelope region (Figure 3). The ECP numbers were counted for HCV strains in each clade. There was no statistical difference of average ECP numbers between clade 2 and clade 3 (data not shown). However, number of ECP were significantly higher than that from clades 2 and 3 in either forward or reverse strand, i.e., forward strand, 6.49 ± 1.99 v.s. 4.91 ± 1.87 ($p = 7.16 \times 10^{-13}$); reverse strand, 7.28 ± 2.04 v.s. 5.58 ± 1.63 ($p = 4.96 \times 10^{-15}$); in combination, 13.77 ± 2.72 v.s. 10.49 ± 2.6 ($p = 1.25 \times 10^{-25}$) (Figure 3).

4. Discussion

The instability of recombinant clones accommodating large or full-length viral genomes is a common concern in RNA virus research, in particular in the generation of infectious viral clones. Potential toxicity of cloned genes is widely accepted as an underlying cause responsible for such instability. In fact, two recent studies have provided solid evidence that cloning instability resulted from the gene expression initiated by cryptic ECP in the insert (Li et al., 2011; Pu et al., 2011). The observation in the present study further strengthens the role of ECP in the cloning stability. In the phylogenetic tree, HCV genotype 1a strains form advanced clustering into three clades (Fan et al., 2009; Pickett et al., 2011). Among all HCV 1a strains done with LRP cloning in the laboratory, easy-to-clone strains can be placed into clade 2 or 3 while difficult-to-clone isolates are located in clade 1 (Figure 2). Corresponding to this finding, the analysis of 395 full-length HCV strains revealed significantly higher ECP numbers in clade 1 in comparison to clades 2 and 3 (Figure 3). In addition, there were no consistent patterns with regard to the ECP distribution along with the HCV genome between clade 1 and clades 2 and 3 (Figure 3). As each ECP on the cloned gene has differential activities and multiple ECPs might interact in synergic manner, the net effect on the cloning stability may be different in cloned genes with similar ECP numbers (Pu et al., 2011). This may explain why some HCV 1a strains, such as #1701 and Liv03, were easily cloned in spite of their high ECP numbers (Table 1). Taken together, for HCV 1a strains, the phylogenetic position may be a more reliable predictor than the simple count of ECP numbers in terms of cloning stability. Since current study included only nine HCV 1a strains, studies with more HCV strains from 1a or other HCV genotypes are necessary prior to the transition of such an implication into a definite conclusion.

There are many possible experimental modifications or protocols that have suggested to enhance cloning stability. One common method is the introduction of intron(s) on viral genomes, which disrupts potential ECP transcription (Yamshchikov et al., 2001). However, this approach only works on individual viral strains and is not feasible for large-scale screening of viral genomes in a defined reverse genetics system. In the current study, almost all reported methods or modifications were tested for HCV LRP cloning. None of them gave satisfactory results (Table 2). The role of 7-deaza-dGTP incorporation on cloning stability was the first report in spite of its extensive application in sequencing and amplification of GC-rich domains (Frey et al., 2008; Broadbent et al., 2010). Although the incorporation of 7-deaza-dGTP may ameliorate or eliminate potential ECPs' activity, it seemed to introduce additional toxicity that actually outweighed its possible benefit in stabilizing cloned inserts.

Whole plasmid mutagenesis generated four positive recombinant clones out of 200 screened colonies. However, the use of vectors from these positive clones did not improve cloning stability (Table 2). Random mutations were found to be scattered on each LRP insert compared with the consensus from 454 sequencing. Thus the stability of these four recombinant clones may be attributable to the altered net effect of multiple ECPs due to the mutations. These mutations may be the adaption during the culture. Alternatively, they may represent authentic mutations from minor variants in a heterogeneous viral population, i.e., the quasispecies (Domingo et al., 2006). In fact, a high positive rate (~100%) was observed in cloning HCV strain sampled at the clinical breakthrough against antiviral therapy, however, the baseline sample from the same patient, i.e., sample Liv23, had only 10% positive rate (Table 1), confirming the role of mutations in cloning stability (Xu et al., 2008).

In summary, all HCV 1a strains tested in the present study have shown remarkable difference in terms of cloning stability, which largely depends on their phylogenetic positions. The HCV 1a strains in clade 1 are very difficult to be cloned compared with those from clades 2 and 3. Putative ECPs on the HCV genome represent an underlying mechanism responsible for such cloning difference. As exhaustive exploration has not found a general approach that can achieve stable cloning for all HCV 1a strains, the selection of appropriate strains, guided by phylogenetic analysis, should be considered prior to attempting to construct infectious HCV 1a clones. In this setting, these observations are not only valuable for establishing HCV 1a cell culture model but also have general implications for other RNA viruses.

Acknowledgments

This work was supported by NIH grants R01 DK80711 (to XF) and R21 AI076834 (to AMD).

References

- Alter MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, Gao F, Moyer LA, Kaslow RA, Margolis HS. Prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N Engl J Med*. 1999; 341:556–562. [PubMed: 10451460]
- Aslanidis C, de Jong PJ. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res*. 1990; 18:6069–6074. [PubMed: 2235490]
- Broadbent S, van der Woude M, Aziz N. Accurate and simple sizing of primer extension products using a non-radioactive approach facilitates identification of transcription initiation sites. *J Microbiol Methods*. 2010; 81:256–258. [PubMed: 20385179]
- Brown WC, Campbell JL. A new cloning vector and expression strategy for genes encoding proteins toxic to *Escherichia coli*. *Gene*. 1993; 127:99–103. [PubMed: 8486292]
- Chambers TJ, Fan X, Droll DA, Hembrador E, Slater T, Nickells MW, Dustin LB, Di Bisceglie AM. E1/E2 quasispecies heterogeneity as a pretreatment variable during pegylated interferon therapy of chronic HCV infection. *J Virol*. 2005; 79:3071–3083. [PubMed: 15709027]
- Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science*. 1989; 244:359–362. [PubMed: 2523562]
- Cirino PC, Mayer KM, Umeno D. Generating mutant libraries using error-prone PCR. *Methods Mol Biol*. 2003; 231:3–9. [PubMed: 12824595]
- Domingo E, Martin V, Perales C, Grande-Perez A, Garcia-Arriaza J, Arias A. Viruses as quasispecies: biological implications. *Curr Top Microbiol Immunol*. 2006; 299:51–82. [PubMed: 16568896]
- Fan X, Mao Q, Zhou D, Xing J, Xu Y, Ray SC, Di Bisceglie AM. High Diversity of Hepatitis C Viral Quasispecies is Associated with Early Virological Response in Patients undergoing Antiviral Therapy. *Hepatology*. 2009; 50:1765–1772. [PubMed: 19937690]
- Fan X, Xu Y, Di Bisceglie AM. Efficient amplification and cloning of near full-length hepatitis C virus genome from clinical samples. *Biochem Biophys Res Commun*. 2006; 346:1163–1172.

- Farci P, Shimoda A, Coiana A, Diaz G, Peddis G, Melpolder JC, Strazzera A, Chien DY, Munoz SJ, Balestrieri A, Purcell RH, Alter HJ. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science*. 2000; 288:339–344. [PubMed: 10764648]
- Frey UH, Bachmann HS, Peters J, Siffert W. PCR-amplification of GC-rich regions: 'slowdown PCR'. *Nat Protoc*. 2008; 3:1312–1317. [PubMed: 18714299]
- Grime SK, Martin RL, Holaway BL. Inhibition of restriction enzyme cleavage of DNA modified with 7-deaza-dGTP. *Nucleic Acids Res*. 1991; 19:2791. [PubMed: 1645867]
- Jung A, Ruckert S, Frank P, Brabletz T, Kirchner T. 7-Deaza-2'-deoxyguanosine allows PCR and sequencing reactions from CpG islands. *Mol Pathol*. 2002; 55:55–57. [PubMed: 11836448]
- Kuiken C, Yusim K, Boykin L, Richardson R. The Los Alamos HCV Sequence Database. *Bioinformatics*. 2005; 21:379–384. [PubMed: 15377502]
- Li D, Aaskov J, Lott WB. Identification of a cryptic prokaryotic promoter within the cDNA encoding the 5' end of dengue virus RNA genome. *PLoS ONE*. 2011; 6:e18197. [PubMed: 21483867]
- Manen D, Goebel T, Caro L. The par region of pSC101 affects plasmid copy number as well as stability. *Mol Microbiol*. 1990; 4:1839–1846. [PubMed: 2082144]
- Nohno T, Kasai Y, Saito T. Cloning and sequencing of the *Escherichia coli* chlEN operon involved in molybdopterin biosynthesis. *J Bacteriol*. 1988; 170:4097–4102. [PubMed: 3045084]
- Nohno T, Saito T, Hong JS. Cloning and complete nucleotide sequence of the *Escherichia coli* glutamine permease operon (glnHPQ). *Mol Gen Genet*. 1986; 205:260–269. [PubMed: 3027504]
- Orosz A, Boros I, Venetianer P. Analysis of the complex transcription termination region of the *Escherichia coli* rrnB gene. *Eur J Biochem*. 1991; 201:653–659. [PubMed: 1718749]
- Pickett BE, Striker R, Lefkowitz EJ. Evidence for separation of HCV subtype 1a into two distinct clades. *J Viral Hepat*. 2011; 18:608–618. [PubMed: 20565573]
- Pu SY, Wu RH, Yang CC, Jao TM, Tsai MH, Wang JC, Lin HM, Chao YS, Yueh A. Successful propagation of flavivirus infectious cDNAs by a novel method to reduce the cryptic bacterial promoter activity of virus genomes. *J Virol*. 2011; 85:2927–2941. [PubMed: 21228244]
- Reese MG. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput Chem*. 2001; 26:51–56. [PubMed: 11765852]
- Ruggli N, Rice CM. Functional cDNA clones of the Flaviviridae: strategies and applications. *Adv Virus Res*. 1999; 53:183–207. [PubMed: 10582099]
- She K. So you want to Work with Giants: The BAC Vector. *BioTeach J*. 2003; 1:69–74.
- Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*. 2005; 41:207–234. [PubMed: 15915565]
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011; 28:2731–2739. [PubMed: 21546353]
- Tellinghuisen TL, Evans MJ, von Hahn T, You S, Rice CM. Studying hepatitis C virus: making the best of a bad virus. *J Virol*. 2007; 81:8853–8867. [PubMed: 17522203]
- Wakita T, Pietschmann T, Kato T, Date T, Miyamoto M, Zhao Z, Murthy K, Habermann A, Krausslich HG, Mizokami M, Bartenschlager R, Liang TJ. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med*. 2005; 11:791–796. [PubMed: 15951748]
- Weiss J, Zucht HD, Forssmann WG. Amplification of gene fragments with very high G/C content: c7dGTP and the problem of visualizing the amplification products. *PCR Methods Appl*. 1994; 4:124–125. [PubMed: 7580884]
- Wild J, Hradecna Z, Szybalski W. Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genome Res*. 2002; 12:1434–1444. [PubMed: 12213781]
- Xu Z, Fan X, Xu Y, Di Bisceglie AM. Comparative Analysis of Near Full-length Hepatitis C Viral Quasispecies in Patients Experiencing Viral Breakthrough during Antiviral Therapy. *J Virol*. 2008; 82:9417–9424. [PubMed: 18667493]
- Yamshchikov V, Mishin V, Cominelli F. A new strategy in design of +RNA virus infectious clones enabling their stable propagation in *E. coli*. *Virology*. 2001; 281:272–280. [PubMed: 11277699]

- Zhang Y, Taiming L, Liu J. Low temperature and glucose enhanced T7 RNA polymerase-based plasmid stability for increasing expression of glucagon-like peptide-2 in *Escherichia coli*. *Protein Expr Purif*. 2003; 29:132–139. [PubMed: 12729734]
- Zhong J, Gastaminza P, Cheng G, Kapadia S, Kato T, Burton DR, Wieland SF, Uprichard SL, Wakita T, Chisari FV. Robust hepatitis C virus infection in vitro. *P Natl Acad Sci USA*. 2005; 102:9294–9299.
- Zhou D, Fan X, Tan D, Xu Y, Tavis JE, Di Bisceglie AM. Separation of near full-length hepatitis C virus quasispecies variants from a complex population. *J Virological Methods*. 2007; 141:220–224.

Highlights

- Instability of recombinant clones harboring large or full-length viral genomes is a common but difficult issue in virological research;
- Exhaustive experimentation that estimated nearly all available options to enhance the cloning stability.
- The first report to reveal remarkable difference with regard to cloning instability among phylogenetic clades of hepatitis C virus genotype 1a.

rrnB T1: ATAAAACGAAAGGCCAGTCTTTCGACTGAGCCTTTCGTTTTAT
 rrnB T2: AAAAGGCCATCCGTCAAGATGGCCTTCT
 tHP: GGTACCCGATAAAAGCGGCTTCCTGACAGGAGGCCGTTTGTGTTTGCAGCCACCTC

Figure 1.
 Schematic representation of the plasmids pCloneKR and pCloneKT. Transcription terminator sequences are shown on the top.

Figure 2.

The Neighbor-Joining tree constructed based on 402-bp partial HCV envelope sequence spanning the hypervariable region 1 (HVR1). The tree included 157 HCV consensus sequences from previous studies (Chamer et al., 2005; Fan et al., 2009) and 26 reference sequences from other HCV genotypes. In contrast to HCV 1b that showed a single star-like structure, HCV 1a strains were clustered into three clades. All HCV strains used for LRP cloning were marked for their positions in clades with colors blue, fuchsia and red corresponding with high, moderate and low clonal stability, respectively.

Figure 3.

Analysis of putative ECP among 395 HCV genotype 1a strains. All HCV 1a strains were clustered into three clades in the Neighbor-Joining tree (A). The ECP numbers (score>0.9) were computed using neural network promoter program on the HCV genome from position 1 to 9035, which was available for all 395 HCV 1a strains (Reese et al., 2001). The number of ECP in clade 1 was significantly higher than that from both clades 2 and 3 (B). In addition, there were differential patterns regarding the ECP frequency distribution along with the HCV genome between clade 1 (n=271) and clades 2 and 3 (n=124) (C).

A summary of HCV samples used for LRP cloning. The HCV genotype, viral load and quasispecies (QS) diversity were determined in previous studies (Chamber et al., 2005; Fan et al., 2009). HCV RNA level was measured by the Roche Amplicor assay and QS diversity was measured by genetic distance (d). The QS diversity for sample H77 was estimated from a published study (Farci et al., 2000). The genetic distance was calculated between each strain and HCV 1a prototype H77 (AF009606). The positive rates of recombinant clones were calculated based on the results of colony PCR/enzyme digestion of thirty randomly picked clones. Gen., genotype; NA, not analyzed. The LRP amplicon from this sample was not sequenced.

Table 1

HCV sample	Gen.	Clade	Viral load (copies/mL)	QS (d)	Genetic distance	ECP number	Positive rate (%)	Reference
1106	1a	1	883000	0.0017	0.071	17	6.7	This study
4701	1a	1	1370000	0.0244	0.048	15	0	This study
H77	1a	1	46670000	0.05	0.002	14	0	This study
1701	1a	2	2370000	0.6865	0.073	16	60	This study
1709	1a	3	4870000	0.5835	0.086	8	70	This study
Liv19	1a	2	964000	0.0277	NA	NA	70	Fan et al., 2006
Liv23	1a	1	572000	0.2611	0.050	15	10	Zhou et al., 2007
Liv02	1a	3	168000	0.0025	0.118	9	75	Xu et al., 2008
Liv03	1a	2	215000	0.3038	0.082	15	35	Xu et al., 2008

Table 2

A brief summary of major approaches to explore the cloning stability. The 9022-bp fragments from samples H77 and #4701, representing difficultly cloned HCV strains (Table 1) were used as the model template. Result is shown as the positive recombinant clones (with 9022-bp insert) among thirty randomly picked clones. App., approach; trans., transcription; NP, not performed due to the negative result from partial plasmid mutagenesis.

App.	Description	Vector	Cell	Result
1	Lowing culture temperature	pClone (3030 bp)	DH10B	0/30
2	Glucose in culture	pClone (3030 bp)	DH10B	0/30
3	Switch insert direction	pClone (3030 bp)	DH10B	0/30
4	LRP purification by gelase	pClone (3030 bp)	DH10B	0/30
5	Mini-vector: pCloneK	pCloneK (2222 bp)	DH10B	0/30
6	Mini-vector: pCloneA	pCloneA (1946 bp)	DH10B	0/30
7	CopyCutter EPI400 E. coli cells	pCloneK (2222 bp)	EPI400	0/30
8	SURE E. coli cells	pCloneA (1946 bp)	SURE	1/30
9	Other E. coli cells	pCloneK (2222 bp)	other	0/30
10	Trans. terminator: rmbT1T2/rmbT1T2	pCloneKR (2565 bp)	DH10B	0/30
11	Trans. terminator: tHP/rmbT1T2	pCloneKT (2419 bp)	DH10B	0/30
12	Inserting par element onto pCloneKR	pCloneKR_Par	DH10B	0/30
13	Inserting par element onto pCloneKT	pCloneKT_Par	DN10B	0/30
14	BAC vector (Epicentre pCC1 system)	pCC1 (8139 bp)	EPI300	1~3/30
15	BAC vector (Lucigen pEZ system)	pEZ (7256 bp)	10G	1~3/30
16	Ligation independent cloning (LIC)	pCloneK (2222 bp)	DH10B	0/30
17	7-deaza-dGTP in LRP amplification	pCloneK (2222 bp)	DH10B	0/30
18	Partial plasmid mutagenesis	pCloneK mutants	NP	NP
19	Whole plasmid mutagenesis	pCloneK mutants	DH10B	0/30

Table 3

The list of primers for sequencing the LRP insert of sample H77 in four stable recombinant clones resulted from the whole plasmid mutagenesis. Primer numbering is according to the HCV H77 strain (GenBank accession no. AF009606). Primer CloneF is from plasmid pCloneK. Primers were designed using the software program Eugene (version 1.01).

Primers	Polarity	Sequence (5'-3')	Position
CloneF	Sense	gcctctatcgcccttcttg	pCloneK
H77_1	Sense	aatgagggttcggtgg	465~482
H77_2	Sense	gcgggtctgtcttctgttg	1045~1065
H77_3	Sense	gatcaacaccaacggcag	1442~1459
H77_4	Sense	gatgtacgtggagggtc	2093~2111
H77_5	Sense	tcctcgtgttcttctgttg	2506~2526
H77_6	Sense	aaccatctcaccctcttcg	3053~3072
H77_7	Sense	ctcattgacaccctgcacc	3556~3574
H77_8	Sense	gctacaagggttggtgctc	3945~3964
H77_9	Sense	cgatgtgtcgtcgtgtc	4491~4508
H77_10	Sense	gggagaacttcttacctgg	4940~4960
H77_11	Sense	atgtggaattcatcagtg	5494~5513
H77_12	Sense	gaaccatgtttcccccac	5979~5996
H77_13	Sense	ggccctgtactcccttc	6403~6421
H77_14	Sense	gaggaggatgagcgggag	6985~7002
H77_15	Sense	tcgtgtcgtcctcaatgc	7450~7468
H77_16	Sense	taagccagctcgtctcatcg	7922~7941
H77_17	Sense	taacatcatgctcctccaacg	8547~8567