ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

A meta-analysis of genotyping methods and nucleotide variation in Hepatitis C Virus

Raju Nagarapu[1], Syed Rahamathulla[1], Sandeep Kumar Vishwakarma[1], Srivatsan Kunnavakkam Vinjimuri[1], Ramachandra Rao VV[1], Pathan K Khanum[1], Mohammed Aejaz Habeeb[1], Vishnu Priya Satti[2], Mohammed Nanne Khaja[1], Aleem Ahmed Khan[1*]

1. Center for Liver Research and Diagnostics, Deccan College of Medical Sciences and Allied Hospitals, Kanchanbagh, Hyderabad-500 058, Andhra Pradesh, INDIA.

2. Department of Genetics, Osmania University, Hyderabad-500 007, Andhra Pradesh, INDIA.

Manuscript Info

Manuscript History:

Received: 14 September 2013
Final Accepted: 22 September 2013
Published Online: October 2013

Key words:
HCV; genotyping methods; nucleotide variations; meta-analysis

Corresponding Author

Abstract

Background: Worldwide distribution of different Hepatitis C Virus (HCV) genotype is responsible for many chronic and severe infections. Nucleotide sequence comparison of HCV genotypes has revealed significant genetic heterogeneity of the HCV genome. Various studies have been reported in order to know sequence heterogeneity in HCV genome to find more appropriate drug for the treatment of HCV infection in more defined manner.

Methods: A search was performed on NCBI for HCV genotyping methods and sequence analyses of 104 peer-reviewed references were used to identify against four highly conserved regions of HCV genome 5'untranslated region (5'UTR), envelope region 1 (E1), CORE, and non structural 5 b region (NS5B). Multiple sequence alignment was performed for 19 subtypes of 6 HCV genotypes by Vector NTI software for 341bp, 573bp, 576bp, and 1777bp amplicons of 5'UTR, CORE, E1, NS5B regions respectively.

Results: The studies performed since 1993-2012 to identify the correct HCV genotype based on four conserved regions 5'UTR, CORE, E1 and NS5B using various techniques were found to be ≥35% in Line Probe Assay (LiPA), ≥22% in PCR-Sequencing, ≥17% in RFLP, 11% type specific PCR, 7% multiplex PCR, 4% nested PCR and 1% in Support Vector Machine (SVM), Primer specific and mis-pair extension analysis (PSMEA), liquid microarray and mass spectroscopy. The study reveals that 5'UTR is the most conserved region along the HCV genome distribution and widely used in HCV genotyping and having conserved variation, CORE has been used for HCV subtype discrimination.

Conclusion: Inno-LiPA is the most widely used method for HCV genotyping and 5'UTR is most conserved region for HCV genotyping while CORE is used for subtype discrimination. The present study demonstrates that all the genotyping methods have been mostly targeted to 5'UTR, CORE, E1 and NS5B showing 5'UTR as most conserved region for genotyping and CORE for HCV subtype discrimination.

1. Introduction

Hepatitis C Virus (HCV) is a positive enveloped single stranded RNA virus of ~9.6kb in length. It belongs to *hepacivirus flaviviridae* family and is responsible for acute as well as chronic illness in human (Choo *et al.*, 1991). The treatment of HCV infection depends on the genotype variability, whole genome sequence and protein products. Distribution of different genotypes worldwide is responsible for many chronic and severe infections. Sequences comparison of genotypes has revealed significant genetic heterogeneity of the HCV genome. Based on the sequence

variability both in coding and non-coding regions, several classification systems have been proposed According to Enomoto et al., (1990) HCV has concluded into two major types K1 and K2. Later Mori, Simmonds, Stuyver, and Bukh reported several additional genotypes and proposed their own classification and nomenclature schemes (Mori *et al.,* 1992; Simmonds *et al.,* 1993; Stuyver *et al.,* 1993, and Bukh *et al.,* 1993, 1994). Since then a number of HCV genotyping systems have been developed.

Several authors performed HCV genotyping by genotyped HCV both by restriction fragment length polymorphism (RFLP) (Nakao *et al.,* 1991 and McOmish *et al.,* 1993) and PCR with genotype-specific primers (Okamoto *et al.,* 1992 and Chayama *et al.,* 1993). On the basis of structural similarities and differences in the genome sequence of HCV has been categorized into six genotypes 1, 2, 3, 4, 5, 6 and moreover 120 subtypes (Bukh *et al.,* 1993, 1994; Maertens and Stuyver, 1997; Simmonds, 1995). To study these genotypes, we selected for genotyping based on variable regions such as the E1, CORE, and NS5B provide enough resolution to reliably determine types and subtypes in HCV genome. However, the 5'UTR is too conserved for accurate discrimination of all genotypes. Nevertheless, the conserved nature of the 5'UTR has made it the preferred target for pan genomic HCV RNA detection tests, and sequence analysis of amplicons, the most efficient way to genotype HCV in a clinical laboratory setting since both tests can be completed with the product from a single amplification reaction.

5'UTR contains 341bp and is the most conserved region of HCV RNA in terms of primary sequence and secondary structures (Nizar *et al.,* 2000). The fact that the structured domain is located at the 5' end of the genome, it stands to reason that they play important roles in viral RNA translation and/or replication (Stephane *et al.,* 2006). The detection of HCV RNA by reverse transcriptase PCR (RT-PCR) has become essential for the diagnosis of HCV infection and for the selection of patients before therapy. The main advantages of RT-PCR include early diagnosis after acute infection and detection of viremia in selected patients. The sensitivity of PCR for HCV RNA detection may vary according to the choice of primers and the handling of pre-extraction samples. Most laboratories use primers specific for the 5' UTR of the HCV genome because it represents the most highly conserved region among HCV genotypes.

CORE region contains 573bp and is very useful for the differentiation of subtypes. This region contains higher mutation than the 5' UTR. The HCV CORE protein has been reported to have many functions. With respect to the virus, the main function of the CORE protein is to form the capsid shell that house and protect the HCV genomic RNA while the virus passes from one cell to another or from one person to another person. However, the HCV CORE protein also modulates many different host pathways by interacting with a variety of cellular factors (Stephane *et al.,* 2006).

The two HCV envelope glycoproteins, E1 and E2 are released from HCV polyprotein by signal peptidase cleavages. E1 region contains 576bp long sequence and it's essential for virus entry into the host cells. The trans-membrane domains of HCV envelope glycoproteins play a major role in E1-E2 hetero dimer assembly and sub-cellular localization. The envelope glycoprotein complex E1-E2 has been proposed to be essential for HCV entry. The glycoproteins present in their envelope are involved in the receptor-binding step (Stephane *et al.,* 2006).

NS5B region contains 1777bpand the sequence motifs highly conserved among all the known RNA-dependent RNA polymerases (RdRps) (Poch *et al.,* 1989). RdRp is considered an important target for drug development (Beaulieu and Tsantrizoa, 2004; Wu and Hong, 2003), as well as useful region for the demonstration of HCV genotypes and subtypes discrimination.

2. HCV Genotyping Methods

Since last few decades a variety of methods have been used for HCV genotype, such as direct nucleic acid sequencing (Simmonds *et al.,* 1993); reverse hybridization line probe assay (LiPA) (Stuyver *et al.,* 1996), subtype-specific reverse transcription (RT)-PCR (Okamoto *et al.,* 1992), DNA RFLP (Das *et al.,* 1993), heteroduplex mobility analysis (Delwart *et al.,* 1993), primer-specific and mispair extension analysis (Hu *et al.,* 1999). Among all these available assays only few methods have been used widely giving more information with more precise and specific way. Most commercial methods target the 5'UTR, which is the region of choice for qualitative and quantitative HCV RNA detection because of its high degree of conservation among different subtypes. The 5'UTR however does not allow sufficient discrimination between closely related subtypes within the genotype and after we needed to develop various methods have been proposed for the discriminate of HCV genotypes, such as RFLP, which utilizes restriction enzymes to recognize genotype-specific cleavage sites in a PCR-amplified DNA fragment, PCR amplification with type-specific primers, however sensitivity and specificity parameters are require for evolution of these method. DNA hybridization technologies, including LiPA are commercially available method to determine HCV genotypes that may have more potential.

2.1 *Inno-Line Probe Assay*

The VERSANT HCV genotype assay (LiPA) is one of the most widely used methods for HCV genotyping. In this assay the 5'UTR of HCV is amplified with biotinylated primers, after which the PCR product is hybridized to a membrane impregnated with genotype-specific probes and detected with streptavidin linked to a colorimetric detector (Stuyver *et al.,* 1996). The Inno-LiPA HCV typing is based on the reverse hybridization principle. Oligonucleotide probes, specific for each subtype, are hybridized with an amplified biotinylated sample. Following hybridization and a washing step, streptavidin labelled with alkaline phosphatase is used to detect the biotinylated hybrid. Color development is chieved by addition of the substrate nitroblue tetrazolium–5-bromo-4-chloro-3-indolylphosphate toluidinium (NBT-BCIP). Hybridization typing based on genotyping methods represent an attractive genotyping option compared to sequencing methods. Incontestably, INNO-LiPA HCV 2.0 demonstrates better performance than INNO-LiPA 1.0, especially for the subtyping of genotype 1 specific samples and the characterization of genotype 6 (Noppornpanth *et al.,* 2006). LiPA allows the easy and rapid determination of HCV genotypes but fails to determine the subtypes of isolates and misclassifies a significant percentage of type 1a isolates as type 1b.

2.2 *PCR-Sequencing*

For several years HCV genotyping using direct sequencing of the DNA have devised an algorithm of sequence analysis for genotypic determination on the basis of sequence analysis in conserved regions. Direct sequence analysis is regarded as the most accurate and specific method for HCV genotyping. Its clinical applicability is limited by its complex protocols, high cost and inability to detect mixed genotypes, although sequence analysis is considered the "gold standard" for HCV genotype determination. It is expensive, time-consuming, and inconvenient for routine use (Zhenyu *et al.,* 2002). However, in clinical practice, it is only truly necessary to distinguish between different genotypes to make more convenient therapeutic decisions (Mondelli and Silini, 1999).

2.3 *PCR-RFLP*

RFLP exploits variations in homologous DNA sequences referring to a difference between samples of homologous DNA molecules that come from differing locations of restriction enzyme sites. The DNA sample is digesting into small DNA fragments by restriction enzymes and the resulting restriction fragments are separated according to their size on gel electrophoresis. RFLP analysis was the first DNA profiling technique inexpensive enough to see widespread application. In addition to genetic fingerprinting, RFLP have also been used for genetic mapping, localization of genes for genetic disorders, determination of risk for disease, and paternity testing.

RT-PCR followed by RFLP has often used for HCV-genotyping and their subtypes. Several studies have demonstrated this conventional technique for the identification for discrimination of correct genotype in many populations by amplifying HCV cDNA and digesting them with correct restriction enzymes such as MboI, AccI and BstNI etc. Because of this various conserved regions of HCV genome has been amplified and subjected to RFLP. Among these the 5'UTR has been frequently used for amplification followed by digestion basis different sets of restriction enzymes to detect different subtypes. However, like in other assays it is not possible to detect all subtypes by using 5'UTR (Lole *et al.,* 2003); attempts have been made to include the conserved part of other sub genomic regions also both in PCR-RFLP as well as RT-PCR (Nakatani *et al.,* 2010). And developed determining HCV sub-types using HCV CORE region RFLP.

2.4 Type-specific PCR

The first type-specific genotyping assay was described by Okamoto *et al.,* (1992). The assay utilized a first-round amplification of a large section of the CORE region, followed by a set of second nested amplification reactions, each using an identical 5'primer but one of four subtype-specific primers (for genotypes 1a, 1b, 2a, and 2b). This method detectes four different-sized amplicons by gel electrophoresis (Okamoto *et al.,* 1992). A similar method was described by Chayama in the NS5 region (Chayama *et al.,* 1993).

Among the different parts of the HCV genome, consisting of three structural, six nonstructural, and two untranslated regions, the CORE gene appears to be most suitable for the genotyping by PCR (Grakoui *et al.,* 1993; Houghton *et al.,* 1991). Genotyping by selective amplification of other parts of the HCV genome, such as the NS5B region, has been undertaken (Chayama *et al.,* 1993). However, genotype-specific primers with sequences deduced from the sequence of this region are less efficient than CORE region primers (Andonov *et al.,* 1994). Furthermore, because of sequence variations in the NS5B region which are much greater than those in the CORE region, it would be difficult to design NS5B primers which can anneal with HCV cDNA of specified genotypes with high degrees of sensitivity and specificity (Miyakawa *et al.,* 1995). Technically, it surfers from various limitations to be a significant tool for HCV genotyping.

2.5 *Nested PCR*

Nested PCR is intended to reduce non-specific binding in products due to the amplification of unexpected primer binding sites. It involves two sets of primers, used in two successive runs of PCR, the second set intended to amplify a secondary target within the first run product and specify the target. The method has been used to identify the

genotypes by targeting the various conserved regions of HCV genome regions such as 5'UTR, NSB etc. It has a similar level of accuracy as the INNO-LiPA method but is simpler to use and significantly less expensive, alternative for HCV genotyping and is capable of reliable genotyping of HCV RNA directly from clinical samples. Therefore, this assay is being used in routine diagnostic laboratories for HCV genotype testing (Mohamed *et al.,* 2012). In recent years it has proved a tremendous tool for HCV genotype analysis.

2.6 ***Support Vector Machine (SVM)***

A learning algorithm which form a set of positively and negatively labeled training vectors that can be used to classify new unlabeled test samples is referred as SVM. It learns the classifier by mapping the input training samples (x1, . . . , xn) into a possibly high-dimensional feature space and seeking a hyperplane in this space which separates the two types of examples with the largest possible margin, i.e. distance to the nearest poi base pairs. If the training set is not linearly separable, SVM finds a hyperplane, which optimizes a trade-off between good classification and large margin (Cristianini *et al.,* 2000). In addition to linear versions of SVMs, they have been extended to nonlinear cases via kernels.

Random forest is a classification algorithm developed by Leo Breiman (2001) that uses an ensemble of classification trees. A forest contains many decision trees, each of which is constructed with instances having randomly sampled features. The prediction is done by a majority vote of decision trees. Random forest uses both bagging (bootstrap aggregation), a successful approach for combining unstable learners, and random variable selection for tree building. Each tree is un-pruned (grown fully), so as to obtain low bias trees; at the same time, bagging and random variable selection result in low correlation of the individual trees. The algorithm yields an ensemble that can achieve both low bias and low variance (from averaging over a large ensemble of low-bias, high-variance but low correlation trees).

2.7 ***Primer specific and mis-pair extension analysis (PSMEA)***

The original assay, S-PSMEA (Hu *et al.,* 1999), was developed by substituting the 32P-labeled primer with a 5'-end Cy 5.5 dye-labeled primer, allowing the assay to be performed with an automated DNA sequencer to detect primer extension. This improvement has led to the development of the assay as a semi automated system. Sensitivity and reliability of S-PSMEA for detection of HCV mixed genotype infections by comparing four major currently available genotyping assays, type-specific PCR (T-S PCR), RFLP analysis, line probe assay (LiPA), and direct DNA sequencing with S-PSMEA and population-based DNA sequencing.

2.8 ***Heteroduplex mobility analysis***

Heteroduplex mobility analysis (HMA) relies on the formation of mismatches when two divergent DNA molecules are mixed, denatured, and allowed to reanneal. This process results in the formation of homoduplexes and heteroduplexes that migrate at different speeds on polyacrylamide gel electrophoresis (PAGE). The mismatches reduce the mobility of the heteroduplexes, which are retarded approximately in proportion to the divergence between the two sequences. Unpaired nucleotides produce larger shifts than mismatched nucleotides (Delwart *et al.,* 1993).

Genotyping by HMA involves mixing a PCR product of unknown genotype separately with a panel of reference products of each genotype and separating the resultant heteroduplexes by PAGE. Ideally, the sequences of the subtypes in the panel should adhere as closely as possible to the consensus sequence of each subtype. Genotype determination relies on the identification of heterologous genotypes in lanes that contain heteroduplexes with reduced mobility. HMA applications have included assessment of quasispecies in human immunodeficiency virus (Delwart *et al.,* 1993) and HCV (Polyak *et al.,* 1998; Sullivan *et al.,* 1998; Wilson *et al.,* 1995) and screening of Influenza B Virus variants (Zou *et al.,* 1998).

2.9 ***Liquid Microarray***

Liquid microarray technology has been used for the detection of many different pathogens (Dumonceaux *et al.,* 2009; Wang *et al.,* 2009; Etienne *et al.,* 2009). One of its key properties is an extensive multiplexing capacity, making it possible to detect different nucleic acid targets simultaneously. This is particularly important for HCV genotyping, as there are many different HCV genotypes and subtypes (Simmonds *et al.,* 1994). HCV genotyping on the basis of two different viral genome regions has the advantage that the results for one region confirm those for the other, from this should also method for possible to detect rare cases of inter-genotypic recombination (Kalinina *et al.,* 2002; Colina *et al.,* 2004). The liquid microarray assay described here had excellent clinical sensitivity (95%) for genotype determination. RNA instability and secondary structures have made it difficult to develop molecular tests for HCV (Duarte *et al.,* 2010).

2.10 ***Pyrosequencing***

Pyrosequencing is a novel experimental technique for determining the sequence of nucleotides in a genome (Fakhrai-Rad *et al.,* 2002, Margulies *et al.,* 2005). The method is faster, less laborious, and cheaper than existing

technologies, but pyrosequencing reads are also significantly shorter and more error-prone (~100–250bp and 5–10 errors/kb) than those obtained from Sanger sequencing (~1000bp and 0.01 errors/kb) (Malet *et al.*, 2003, Huse *et al.*, 2007, Wang *et al.*, 2007). Pyrosequencing of a virus population produces many reads, each of which originates from exactly one but unknown haplotype in the population. Thus, the central problem is to reconstruct from the read data the set of possible haplotypes that is consistent with the observed reads and to infer the structure of the population, i.e., the relative frequency of each haplotype.

The alignment step of the proposed procedure is straight forward for the data analyzed here and has been discussed elsewhere (Wang *et al.*, 2007). Due to the presence of a reference genome, only pair-wise alignment is necessary between each read and the reference genome.

3. Targeted Data Entry

The complete genotype articles were retrieved in Dec, 2012 by performing queries from the NCBI homepage search interface (http://www.ncbi.nlm.nih.gov/) by specifying ‘‘HCV Genotyping’’, and also searched from other related journals and articles. The NCBI data are derived from the peer-reviewed literature indexed in Pubmed. From the literatures, we have focused on genotyping methods for HCV used in various countries since 1993-2012. The data were searched for four regions of HCV 5’UTR, CORE, E1, NS5B for genotype analysis. Most of the articles were found to support 5’UTR as the most conserved region of HCV genome and is most interested region for the authors for HCV genotyping analysis.

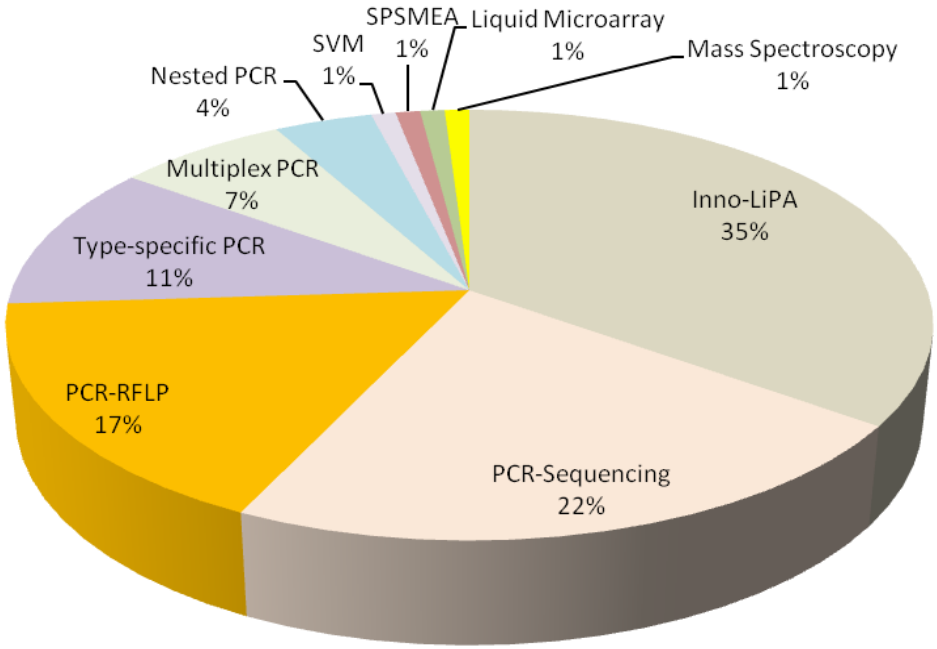
4. Results

A search was done on NCBI for HCV genotyping methods and sequence analysis of 104 peer-reviewed references for genotyping methods based on various highly conserved regions of HCV genome. These references described a total of ten types of genotyping methods of various regions and 19 full length genome sequences of HCV.

4.1 *HCV genotyping*

The studies performed since 1993-2012 to identify the correct HCV genotype based on four conserved regions 5’UTR, CORE, E1 and NS5B using various techniques was found to be ≥35% in Line Probe Assay (LiPA), ≥22% in PCR-Sequencing, ≥17% in RFLP, 11% type specific PCR, 7% multiplex PCR, 4% nested PCR and 1% in SVM, SPSMEA, liquid microarray and mass spectroscopy each (Fig 1).

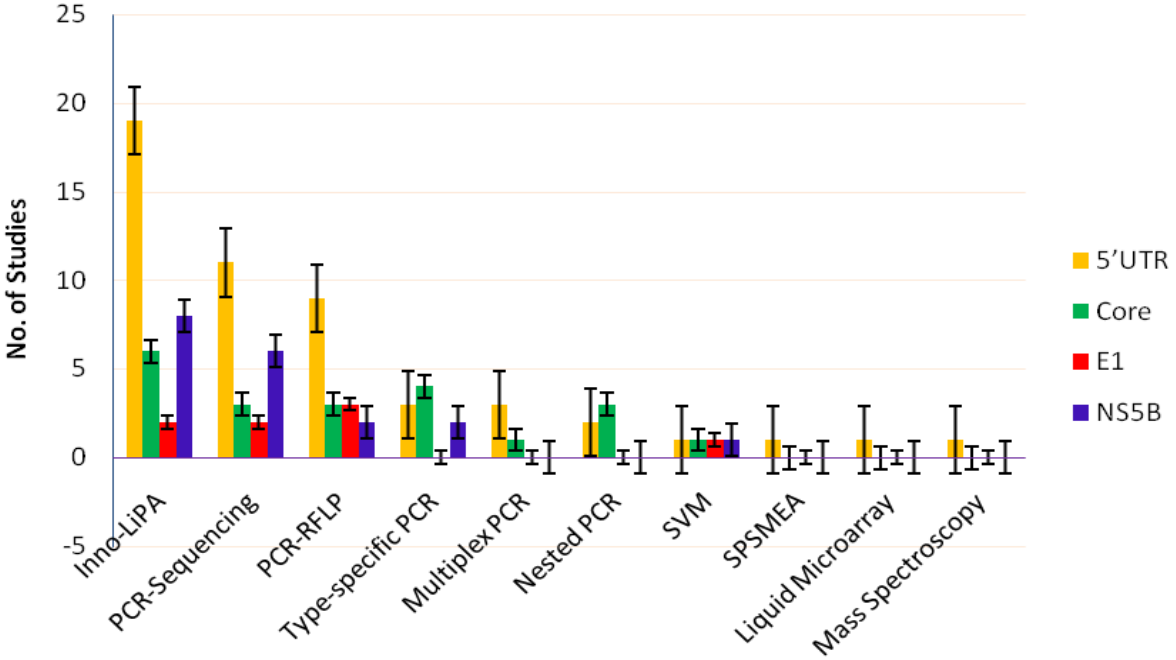
Fig 1: Global studies percentage for HCV genotyping based on 5’UTR, CORE, E1 and NS5B regions using various methods.



In these assays most of the studies ware found in the 5’UTR, CORE and NS5B regions using common methods of genotyping such as LiPA, PCR-sequencing and PCR-RFLP whereas it was highest in RFLP when compared to E1.

The studies for other methods were significantly very less and some were found to be negligible in HCV genotyping such as SPSMEA, liquid microarray and mass spectroscopy (Fig 2).

Fig 2: Genotyping methods used for the distribution of 5'UTR, CORE, E1 and NS5B regions from 1993-2012 (Percentage error bars are represented with 5% value).



4.2 *Nucleotide variation in HCV genotypes and subtypes*

To identify the sequence variability, 19 full length HCV genome sequences were downloaded named 1a, 1b, 1c, 2a, 2b, 2c, 2i, 2k, 3a, 3b, 3k, 4a, 5a, 6a, 6b, 6d, 6g, 6h and 6k from the NCBI and compared the data for four specified regions. Sequence alignment of 19 subtypes of all 6 genotypes of HCV highest sequence variability was found in NS5B region and least was in 5'UTR. 341bp, 573bp, 576bp, and 1777bp for 5'UTR, CORE, E1, NS5B regions respectively were subjected to alignment for the analysis of sequence variability in HCV genotypes and subtypes. Maximum difference was observed in envelope gene E1 when compared to other three regions used in the study (Fig 3). NS5B has also shown to have more differences due to enlarged size in comparison to others but percentage of nucleotide differences in genotype to genotype is very low than E1 (Fig 4).

Fig 3: Nucleotide variation in 5'UTR, CORE, E1, NS5B regions of HCV genotypes and subtypes.

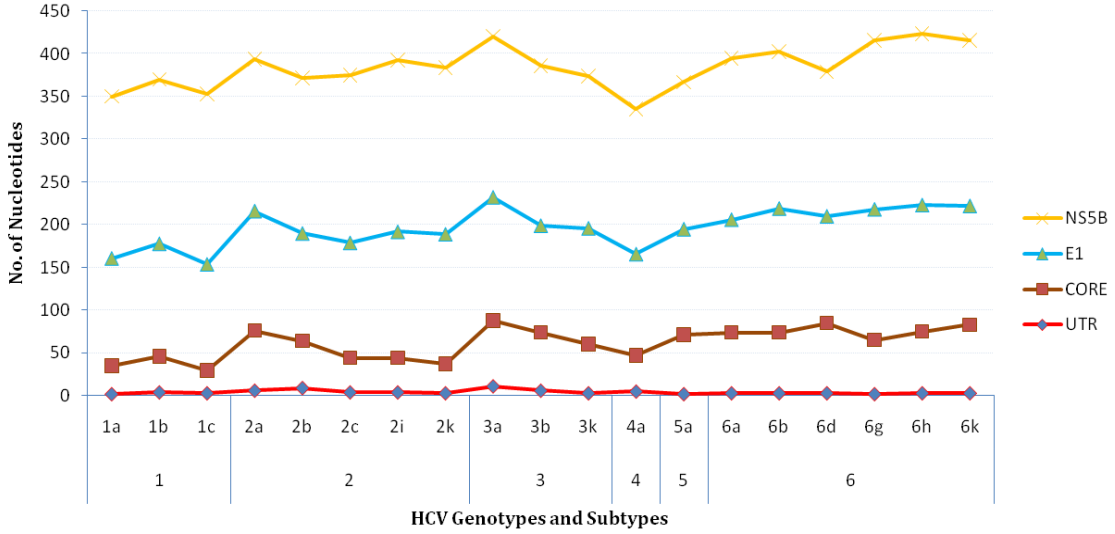
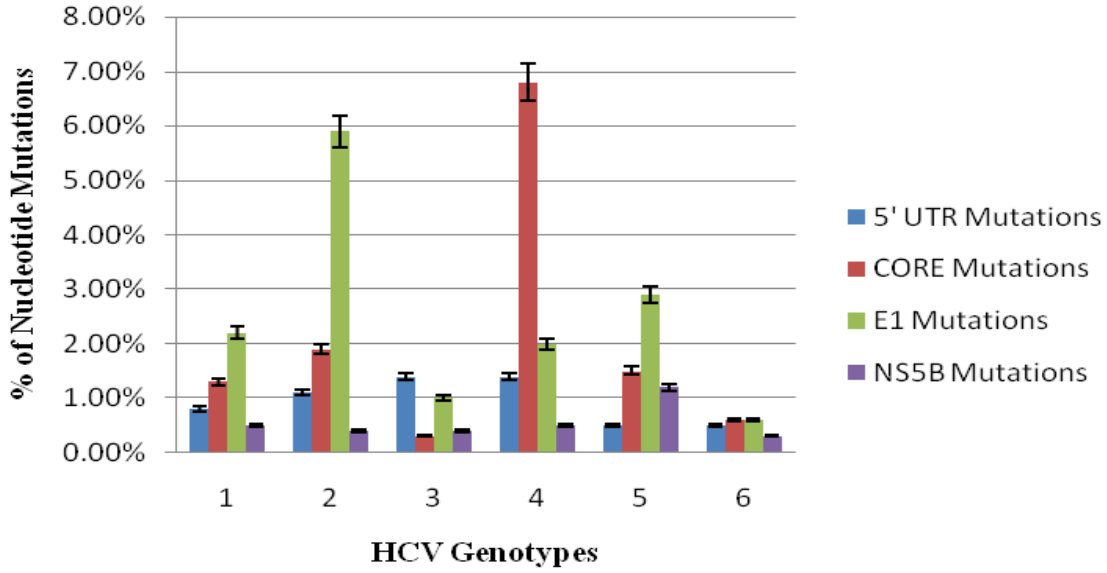


Fig 4: Percentage of nucleotide mutations occurring in varies genotypes of HCV in 5'UTR, CORE, E1 and NS5B (Percentage error bars are represented with 5% value).



4.2.1 ***5'UTR region***

In 5'UTR mutation rate is very low. we analyzed the sequence of HCV 5'UTR region, we get in genotype 1 have only three bp(0.8%) mutations in position that 11, 180 and 224. In genotype 2, four bp(1.1%), genotype 3, five bp(1.4%), genotype 4, five bp(1.4%), genotype 5, only two bp(0.5%), and genotype 6 also have in the difference of two bp(0.5%). The 5'UTR has shown to be inappropriate to discriminate HCV strains, especially at the subtype level. The main failure is misclassification of genotype 6 as genotype 1 due to the same nucleotide mutations, and second thing is lack of sub typing genotypes 2, 3, 4 related to the diversity within this region.

4.2.2 ***HCV*** **CORE** ***region***

HCV CORE region genotype 1 have eight mutations (1.3%), genotype 2, eleven bp(1.9%), genotype 3 only two bp(0.3%), genotype 4 have more in this region thirty nine bp(6.8%), genotype 5, nine bp(1.5%), and genotype 6 also have in the difference of four bp(0.6%) in this region.

4.2.3 ***E1 region:***

The complete nucleotide sequence has been determined of the putative envelope 1 (E1) gene in 19 full length genomes of HCV and observed that in genotype 1 has thirteen nucleotide (2.2%) differences, genotype 2 has more differences than in its 34bp(5.9%), genotype 3, six bp(1.0%), genotype 4, twelve bp(2.0%), genotype 5, seventeen bp(2.9%), and genotype 6 have four nucleotide (0.6%).

4.2.4 NS5B ***region:***

Mutations in NS5B were found very low compared to the other HCV genome sequences. In the study we analyzed 19 full length genomes of HCV and observed that in genotype 1, 9 nucleotide (0.5%) difference, genotype 2, 8bp (0.4%), genotype 3, 8bp (0.4%), genotype 4, 9bp (0.5%), genotype 5, 22bp(1.2%), and genotype 6 have 7 nucleotide (0.3%) mutations are present.

5. Discussion

We report the first meta-analysis of HCV Genotyping methods and nucleotide variation in different genotypes and subtypes in four commonly used regions 5'UTR, CORE, E1 and NS5B. For the identification and classifications of viral isolates, different genotyping methods have been used since last few years and need to overcome from in appropriate methodologies to get better results with high accuracy and efficacy. The present study reveals the most commonly used method for HCV genotyping that will provide an idea to overcome rarely used methods. In this point of view pyro-sequencing gives a better answer for sequence analysis. It also helps to choose the best methodology for the clinical diagnosis of HCV strains in various regions of the world. The correlations of detectable mutations in HCV genome have shown more consequences with regards to viral infectivity, pathogenesis

or response to difficult antiviral agents. Mutation analysis in all HCV genotypes and subtypes will provide the better understanding about the sequence similarity and pathogenesity of the virus.

The present study also demonstrates that all the genotyping methods used have been mostly targeted to 5' UTR, CORE, E1 and NS5B. The analysis has shown that 5' UTR is the most conserved region in HCV genome in respect to all genotypes. some researchers have built a global position weight matrix (PWM) for the HCV genome and found stronger discriminated power in nucleotide sequence signatures of E1 and NS5B differentiating major HCV genotypes and subtypes than that from 5'NCR and CORE regions. The study also resembles the same showing more variability in E1 and NS5B regions in comparison to others. In most of the studies Inno-LiPA was used for the genotyping discrimination due to less cost, time saving and more precise identification of various genotypes and subtypes. Whereas PCR-sequencing has proved to be used less frequently for genotyping as compared to the Inno-LiPA but ease of use and sequence reading quality makes it much better to see the mutation and sequence discrimination. Direct sequencing have also been the most accurate method for HCV genotyping, however, many other methods have been used because of the expense and technical difficulties of direct sequencing.

The most frequently used methods in clinical laboratories are LiPA (line probe assay) and sequencing of the 5'UTR (both from Bayer Diagnostics). Both assays require two steps: generation of the PCR amplicon and then evaluation of the amplicon by either hybridization or sequencing. An additional disadvantage is the use of the 5'UTR, which is less informative for genotyping than are other, more variable regions of HCV. CORE region have been proved better for HCV subtyping due to subtype specific variation in only few nucleotides than other regions. Whereas type specific PCR is the simplest method for genotyping but time consuming, cost futile and need of more expertise pulls it back as compared to the others. Other methods also suffered from some drawbacks and used in fewer studies. Taken together, the study reveals that Inno-LiPA is most widely used method for HCV genotyping and 5'UTR is most conserved for HCV genotype while CORE is for HCV subtype discrimination. However, there is still need to get the better option for genotyping tools for HCV discrimination with other similar viruses and more data survey is needed to understand its biology and drug response in pathological conditions.

6. Conclusion

Our analysis allowed investigation of HCV genotyping methods related to identify genotypes and subtypes of HCV worldwide. First, comparison of data derived from Pubmed articles revealed genotyping methods, though due to nucleotide variations in HCV genome. Inno-LiPA is most widely used method for HCV genotyping and 5'UTR is most conserved region for HCV genotyping and CORE has been identified for subtype discrimination. The present study demonstrates that all the genotyping methods have been mostly targeted to 5'UTR, CORE, E1 and NS5B showing 5'UTR as most conserved region for genotyping and CORE for discrimination of HCV subtypes.

7. Acknowledgement

We are thankful to Mr. Iqbal for his valuable support in manuscript preparation.

8. Conflict of interest

The authors declare no conflict of interest

9. References

1. Andonov, A., and R. K. Chaudhary. (1994) Genotyping of Canadian hepatitis C virus isolates by PCR. *J. Clin. Microbiol*. 32:2031–2034.
2. Beaulieu, PL., Tsantrizos YS. (2004) Inhibitors of the HCV NS5B polymerase:new hope for the treatment of hepatitis C infections. *Curr Opin Investig Drugs*. *5:*838-50.
3. Breiman L. (2001) Random Forests. *Machine Learning.* 45:5-32.
4. Bukh, J., R. H. Purcell, and R. H. Miller. (1993) At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. *Proc. Natl. Acad. Sci. USA.* 90:8234–8238.
5. Bukh, J., R. H. Purcell, and R. H. Miller. (1994) Sequence analysis of the CORE gene of 14 hepatitis C virus genotypes. *Proc. Natl. Acad. Sci. USA.* 91:8239– 8243.

6. Chayama, K., A. Tsubota, Y. Arase, S. Saitoh, I. Koida, K. Ikeda, T. Matsumoto, M. Kobayashi, S. Iwasaki, S. Koyama, T. Moringa, and H. Ku-mada. (1993) Genotypic subtyping of hepatitis C virus. *J. Gastroenterol. Hepatol.* 8:150–156.
7. Choo QL. (1991) Genetic organization and diversity of the hepatitis C virus. *Proc. Natl. Acad. Sci. U. S. A.* 88:2451–2455.
8. Colina R, Casane D, Vasquez S, Garcia-Aguirre L, Chunga A. (2004) Evidence of intratypic recombination in natural populations of hepatitis C virus. *J Gen Virol.* 85:31–37.
9. Cristianini N, Shawe-Taylor J. (2000) An Introduction to Support Vector Machines. *Cambridge University Press, Cambridge, UK*.
10. Das, M.R., Ali, N., Aruna, B., Rama, D., Nambiar, A., Rehana, Z., Shekhar, M.S., Siddiqui, A., Habibullah, C.M. (1993) Indian strains of Hepatitis C: prevalence and detection. *Curr. Sci.* 65: 477–483.
11. Delwart, EL., EG. Shpaer, J. Louwagie, FE. McCutchan, M. Grez, H. Rubsamen-Waigmann, and JI. Mullins. (1993) Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 env genes. *Science*. 262:1257–1261.
12. Duarte, CAB. Foti L. Nakatani, SM. Riediger, IN. Poersch, CO. (2010) A Novel Hepatitis C Virus Genotyping Method Based on Liquid Microarray. *PLoS ONE* 5(9): e12822. doi:10.1371/journal.pone.0012822.
13. Dumonceaux, TJ. Schellenberg, J. Goleski, V. Hil,l JE. Jaoko, W. (2009) Multiplex detection of bacteria associated with normal microbiota and with bacterial vaginosis in vaginal swabs using oligonucleotide-coupled fluorescent microspheres. *J Clin Microbiol.* 47(12): 4067–4077.
14. Enomoto, N., A. Takada, T. Nakao, and T. Date. (1990) There are two major types of hepatitis C virus in Japan. *Biochem. Biophys. Res. Commun.* 170:1021–1025.
15. Etienne KA, Kano R, Balajee SA, (2009) Development and validation of a microsphere-based Luminex assay for rapid identification of clinically relevant aspergilli. *J Clin Microbiol.* 47(4):1096–1100.
16. Fakhrai-Rad, H. Pourmand, N. Ronaghi, M. (2002) Pyrosequencing: An accurate detection platform for single nucleotide polymorphisms. *Hum Mutat.* 19:479–485.
17. Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice. (1993) Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. *J. Virol*. 67:832–2843.
18. Houghton, M., A. Weiner, J. Han, G. Kuo, and Q. L. Choo. (1991) Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. *Hepatology.* 14:381–388.
19. Hu, Y. W., P. Tranqui, E. Balaskas, V. Scalia, A. Giulivi, P. Gill, S. Cassol, J. B. Angel, C. M. Kovacs, and G. Sher. (1999) A sensitive genotyping assay, primer specific and mispair extension analysis (PSMEA) for detection of low levels of HIV-1 drug resistance mutants. *Can. J. Infect. Dis.* 10(Suppl.):9B.
20. Huse, S. Huber, J. Morrison, H. Sogin, M. Welch, D. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* 8:R143.
21. Kalinina, O. Norder, H. Mukomolov, S. Magnius, LO. (2002) A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J Virol.* 76:4034–4043.
22. Lole, KS. Jha, JA. Shrotri, SP. 0(2003) Comparison of hepatitis C virus genotyping by 5‘ noncoding region and CORE-based reverse transcriptase PCR assay for determining subtype distribution in India. *J Clin Microbiol.* 41:5240–5244.
23. Maertens, G., Stuyver, L., Genotypes and genetic variation of hepatitis. In: Harrison, T.J., Zuckerman, A. (Eds.), (1997) The Molecular Medicine of Viral Hepatitis. *John Wiley and Sons, Chichester, England*, 183–233.
24. Malet, I. Belnard, M. Agut, H. Cahour, A. (2003) From RNA to quasispecies: A DNA polymerase with proofreading activity is highly recommended for accurate assessment of viral diversity. *J Virol Methods.* 109:161–170.
25. Margulies, M. Egholm, M. Altman, WE. Attiya, S. Bader, JS. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* 437:376–380.
26. McOmish, F. Cha, SW. Dow, BC. Gillon, J. Frame, WD. Crawford, RJ. Yap, PL. Follett, EAC. and Simmonds, P. (1993) Detection of three types of hepatitis C virus in blood donors: investigation of type-specific differences in serologic reactivity and rate of alanine aminotransferase abnormalities. *Transfusion.* 33:7–13.
27. Miyakawa, Y. Okamoto, H. and Mayumi, M. (1995) Classifying hepatitis C virus genotypes. *Mol. Med. Today.* 1:20–25.

28. Mohamed Abbas Shemis, Dina Mohamed El-Abd, Dalia Ibrahim Ramadan, Mohamed Ibrahim El-Sayed, Bassem Shenoda Guirgis, Mohamed Ali Saber, Hassan Mohamed El- Said Azzazy. (2012) Evaluation of Multiplex Nested Polymerase Chain Reaction for Routine Hepatitis C Virus Genotyping in Egyptian Patients, *Hepat Mon*. 12(4):265-270.
29. Mondelli, M.U. and Silini, E. (1999) Clinical significance of hepatitis C virus genotypes. *J. Hepatol.* 31:65-70.
30. Mori, S., N. Kato, A. Yagyu, T. Tanaka, Y. Ikeda, B. Petchclai, and P. Chiewslip. (1992) A new type of hepatitis C virus in patients in Thailand. *Biochem. Biophys. Res. Commun.* 183:334–342.
31. Nakao, T., N. Enomoto, N. Takada, and T. Date. (1991) Typing of hepatitis C virus genomes by restriction fragment length polymorphism. *J. Gen. Virol.* 72:2105–2112.
32. Nakatani SM, Santos CA, Riediger IN. (2010) Development of Hepatitis C Virus Genotyping by Real-Time PCR Based on the NS5B Region. *PLoS One.* 5:e10150.
33. Nizar N. Zein. (2000) Clinical Significance of Hepatitis C Virus Genotypes *Clinic. Microbiol Rev.* 223–235.
34. Noppornpanth, S., E. Sablon, K. De Nys, T. X. Lien, J. Brouwer, M. Van Brussel, S. L. Smits, Y. Poovorawan, A. D. Osterhaus, and B. L. Haagmans. (2006) Genotyping hepatitis C viruses from South East Asia using a novel line probe assay that simultaneously detects 5'UTR and CORE regions. *J. Clin. Microbiol.* 44:3969–3974.
35. Okamoto, H. (1992) Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes. *Virology.* 188: 331–341.
36. Okamoto, H., Y. Sugiyama, S. Okada, K. Kurai, Y. Akahane, Y. Sugai, T. Tanaka, K. Sato, F. Tsuda, Y. Miyamura, and M. Mayumi. (1992) Typing hepatitis C virus by polymerase chain reaction with type-specific primers: application to clinical surveys and tracing infectious sources. *J. Gen. Virol.* 73:673–679
37. Ping Qiu, Xiao-Yan Cai, Wei Ding, Qing Zhang, Ellie D Norris and Jonathan R Greene. (2009) HCV genotyping using statistical classification approach *Journal of Biomedical Science*, 16, 62 doi:10.1186/1423-0127-16-62.
38. Poch, O., Sauvaget, I., Delarue, M., and Tordo, N. (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J. 8*, 3867-3874.
39. Polyak, S. J., S. McArdle, S. L. Liu, D. G. Sullivan, M. Chung, W. T. Hofgartner, R. L. Carithers, Jr., B. J. McMahon, J. I. Mullins, L. COREy, and D. R. Gretch. (1998) 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.* 72:4288–4296.
40. Simmonds P, Smith F, McOmish P, Yap J, Kolberg M, (1994) Identification of genotypes of hepatitis C virus by sequence comparisons in the CORE,E and NS-5 Regions. *J Gen Virol.* 75:1053–1061.
41. Simmonds, P., E. C. Holmels, T. A. Cha, S. W. Chan, F. McOmish, B. Irvine, E. Beall, P. L. Yap, J. Kolberg, and M. S. Urdea. (1993) Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic tree analysis. *J. Gen. Virol.* 74:2391–2399.
42. Simmonds, P. (1995) Variability of hepatitis C virus. *Hepatology.* 21:570.
43. Stephane Chevaliez and Jean-Michel Pawlotsky. (2006) Hepatitis C Viruses Genomes and Molecular Biology, *Norfolk (UK): Horizon Bioscience*, ISBN-10: 1-904933-20-3ISBN-13: 978-1-904933-20-5,
44. Stuyver, L., A. Wyseur, W. van Arnhem, F. Hernandez, and G. Maertens. (1996) Second-generation line probe assay for hepatitis C virus genotyping. *J. Clin. Microbiol.* 34:2259–2266.
45. Stuyver, L., R. Rossau, A. Wyseur, M. Duhamel, B. Vanderborght, H. V. Heuverswyn, and G. Maertens. (1993) Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. *J. Gen. Virol.* 244:1093–1102.
46. Sullivan, DG. Wilson, JJ. Carithers, RL. Perkins, JD. and Gretch, DR. (1998) Multigene tracking of hepatitis C virus quasispecies after liver transplantation correlation of genetic diversification in the envelope region with asymptomatic or mild disease patterns. *J. Virol.* 72:10036–10043.
47. Wang, C. Mitsuya, Y. Gharizadeh, B. Ronaghi, M. Shafer, RW. (2007) Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome Res.* 17:1195–1201.
48. Wang, W. Ren, P. Sheng, J. Mardy, S. Yan, H. (2009) Simultaneous detection of respiratory viruses in children with acute respiratory infection using two different multiplex reverse transcription-PCR assays. *Virol Methods.* 40(5):1621–1622.
49. Wilson, JJ. Polyak, SJ. Day, TD. and Gretch, DR. (1995) Characterization of simple and complex hepatitis C virus quasispecies by heteroduplex gel shift analysis: correlation with nucleotide sequencing. *J. Gen. Virol.* 76:1763– 1771.

50. Wu, J. and Hong, Z. (2003) Targeting NS5B RNA-dependent RNA polymerase for anti-HCV chemotherapy. *Curr. Drug Targets Infect Disord. 3*:207-19.
51. Zhenyu, C. and Karen EW. (2002) Hepatitis C Virus Genotyping: Interrogation of the 5' Untranslated Region Cannot Accurately Distinguish Genotypes 1a and 1b *J. Clinic Microbiol.* 3127–3134.
52. Zou, S. Stansfield, C. and Bridge, J. (1998) Identification of new influenza B virus variants by multiplex reverse transcription-PCR and the heteroduplex mobility assay. *J. Clin. Microbiol.* 36:1544–1548.