



Hepatitis C Virus: Computational Biology Approach towards the Research - A Review

Abhishek Soni*

School of Bioengineering and Biosciences, Lovely Professional University, Punjab, India

**Corresponding e-mail: abhisheksonihmr@gmail.com*

ABSTRACT

Chronic Hepatitis C Virus is a common reason for liver disease and it is the common signal for the transplantation of liver in the area of the US, Australia and in European countries. The disease included 3 percent of the total global population caused by the HCV. HCV virus, a common infection caused by blood-borne mostly seen in the US which includes 40% of chronic liver disease. Wide-reaching approx. 171 million of peoples they are infected from chronic Hepatitis C virus and some of those who are chronically infected will form cirrhosis or liver cancer, hepatic failure or hepatocellular carcinoma which leads to thousands of death every year. Despite the study of HCV for the past 15 years, our knowledge towards the infection caused by HCV has been partial by our inability to grow the virus in cell culture. There are some antiviral medicines that can cure some of the HCV infections and it will reduce the effect of death from cancer and cirrhosis but identification and treatment of Hepatitis C infection are very less. The HCV infection rate is relatively high with replication range between 10^{10} to 10^{12} virions, and a half-life of 2-4 hours. The HCV RNA mutates rapidly because of the lack of error proofreading by viral RNA polymerase. And these increases in genotype mutation and their subtypes make the research to develop HCV vaccine a challenge. As the technology is growing so vast and computational biology are one of the technology that are changing the way and methods to understand the viruses, mostly in the field of genome sequencing, epigenetics, evolution, and transcriptomic analysis, whereas NGS (Next-Generation Sequencing) provides a great platform for the researcher to get better quality and quantity in various fields. Now to get the entire genomic variation data it has been now possible for laboratory-based experiments and to investigate genetic variation and its structure, computation based mechanism and analysis of data on genomic level comes with complexity.

Keywords: HCV, Chronic, Liver, Transcriptomic, Epigenetics, Genomic, NGS, Genome sequencing

INTRODUCTION

Hepatitis virus is a small size having 55-65 nm single-stranded RNA virus. HCV is the cause of hepatitis C and cancers such as liver cancer, Hepatocellular carcinoma, and lymphomas in humans. The HCV virus contains hereditary material which is surrounded by RNA and an icosahedral protective shell of protein and further, it is encased in a lipid envelope of cellular origin. Two viral glycoproteins, E1 and E2 are embedded in the lipid envelope (Figure 1). Hepatitis C Virus has a single-stranded RNA genome. The genome consists of a single ORF that is approx. 9600 nucleotide base long (9.6 kb). This single open reading frame is translated to produce smaller active proteins. If we look at the taxonomy of the HCV it belongs to the Flaviviridae family. Further, the Flaviviridae is classified into three genera [1,2]. These are Flavivirus, Pestivirus, and Hepacivirus. Flavivirus genus includes yellow fever, dengue, Jap. Encephalitis and Tick-borne virus. Whereas the Pestivirus includes swine fever, viral diarrhea, and border disease virus. Instead of these, the HCV that has 7 genotypes and various subtypes belongs to the Hepacivirus genera which include GB virus B (GBV-B) and tamarin virus and they are closely towards human virus GB virus [3].

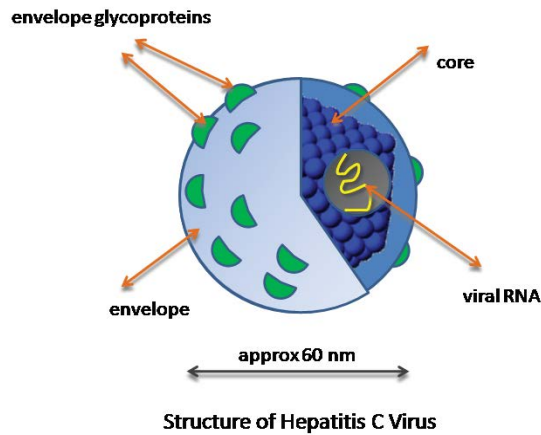


Figure 1 Structure of HCV (Alter, et al.) [1]

As the HCV have various genotype and subtypes. So, here (Table 1) shows the total number of strains and complete genomes that are present in each genotype of the HCV virus [3,4].

Table 1 List of Genotype that is confirmed until 2017 with a total number of strains and their complete Genomes present in each Genotype of HCV virus (Donal, et al.) [5]

Sr. No.	Genotype	Strains	Complete Genomes
1	Genotype 1	65061	1489
2	Genotype 2	3733	637
3	Genotype 3	8975	597
4	Genotype 4	2039	79
5	Genotype 5	762	20
6	Genotype 6	3163	252
7	Genotype 7	9	4

As per the several genotypes that are present in the HCV virus, According to the ViPR (Virus Pathogen Database and Analysis Resource), they have shown the Genome statistics that are discovered till yet (Table 2).

Table 2 Hepatitis C Virus Genome Statistics (ViPR, 2017)

Hepatitis C Virus Genome Statistics	
Strains	149820
3D Protein Structure (PDB)	384
Experimentally Determined Epitopes (IEDB)	7652
Genomes with Clinical Metadata (NIAID GSCID, manual curation)	367
Mature Peptides	348292
Sequence Features with variant Types	367
Protein with Predicted Epitopes	364826
Total Genomes	223309
Complete Genomes	3123
Proteins	5,42,386
Functional Annotation	193811

Worldwide approx. 171 million of peoples they are infected from chronic Hepatitis C virus and some of those who are chronically infected will form cirrhosis or liver cancer, hepatic failure or hepatocellular carcinoma which leads to thousands of death every year. If we look at 2015, the study shows that there were 1.75 million of new HCV infections occur. The treatment of the Hepatitis C virus had changed for the past few years. As per the study in 2016, the FDA

gave approval to a medicine that is a combination of the two components elbasvir and grazoprevir called Zepatier. It has been shown to have the capability to cure the disease in almost 80 - 90%. HCV has been located globally and the most affected region where HCV infection is mostly seen are Eastern Mediterranean, South Africa, and Europe with the existence of 1.5% and 2.3% respectively. And it had been varied from 0.1% to 1.0% in other coastal regions. In today world the use of contaminated or unsafe injections in the sector of health care and injecting drug they are then measured and very rare methods of HCV transmissions.

The core envelope protein E1-E2-p7 on N terminal NS protein having NS2-NS3-NS4A-NS4B- NS5A-NS5B-C terminal. The complete NS fragment of the proteins NS2 – NS5B group depends on the act of viral proteinases. The NS2, NS3 are cleaved by a metal-dependent catalytic proteinase that is encoded with the NS2 and the NS3 region. In the core protein, there are 191 amino acids are present in it. On the other side, on the basis of hydrophobicity, the core regions are divided into three domains (Table 3) [5].

Table 3 Domains and Number of residue and its region

Domain	Number of Residues	Region
Domain 1	01-117	Two Hydrophobic region
Domain 2	118-174	The hydrophobic region at p21 at c terminus
Domain 3	175-191	Signal sequence for E1 protein

E1 protein act as fusogenic and having 4 -5 N linked glycans whereas the E2 protein as a receptor-binding site and has 11 N glycosylation sites. The p7 protein plays a role in virus morphogenesis. The protein is having a 63 amino acid membrane which is located in the region of the endoplasmic reticulum. NS2 protein is a 21-23 kDa protein that has protease activity. NS3 protein which is having a 67 kDa protein and has a serine protease activity on their N terminal whereas in C terminal they have an NTPase and helicase activity. It is situated in the endoplasmic reticulum and they form a heterodimer. The NS5b proteins are having 65 kDa and it is a Serine Protease. NS5B which have the function of replicating the HCV viral [6].

Over the past few years, the HCV virus leads to liver disease and cause cirrhosis. In most of the cases, cirrhosis will lead to liver failure, cancer, esophageal and gastric varices. HCV is caused by blood to blood contacts that are related to the intravenous use of drugs [7,8].

Globally, in a large proportion liver disease and cancer are the diseases that are caused by HCV infection. In sub-Africa areas, the infection is highly seen. Among these West Africa was identified as a geographic origin of HCV genotype. Globally approx.185 million people are infected from HCV infection and approx. 3 lakh people die every year from HCV related liver diseases. The disease varies from asymptomatic infection to cirrhosis and hepatocellular carcinoma [9]. New technologies and therapies with high rates of sustained virological response or cure are impractical owing to exorbitant cost for the majority of HCV infected persons in sub-Africa, where the infection is highly prevalent. In Ghana, the occurrence of infection among blood donors goes beyond endemic levels with up to 11.6% among male donors. The high occurrence of HCV infection between the blood donors and the factors that are responsible for the transmission of HCV virus are sustainable and efficient and these are the major cause of the infection among people. Exposures to infected blood and blood products through unsafe injection practices, transfusion of untested blood, unsterile medical and dental procedures and traditional medical and cosmetic procedures such as scarification are major risk factors for HCV transmission. Recently reported that the major risk factors that are associated with HCV infection among blood donors in Kumasi, Ghana are home birth, tribal scarring, and HBV co-infection [10].

Hepatitis C infection a hepatotropic RNA infection causes dynamic liver harm, which may bring about cirrhosis and carcinoma. Universally in the vicinity of 64 to 103 million entities, they are constantly contaminated or having a chance of HCV. Significant hazard aspects for this is a blood-borne infectious disease that is hazardous infusion tranquilize utilize and unsterile therapeutic systems (iatrogenic contaminations) with high HCV prevalence. The analytic methods that incorporate serum HCV counteracting agent testing, HCV RNA estimation, viral genotype, and subtype assurance and evaluation of protection related substitutions. Different Direct-Acting Antiviral operators (DAAs) have turned out to be accessible and these targets three proteins associated with pivotal strides of the HCV life cycle:

NS3/4A protease, NS3A protein, and the Serine Protease NS3B protein. The blend of a few of these DAAs can cure HCV disease in 90% of the patients which includes common people that have been hard to treat previously. For whatever length of time that a prophylactic immunization is not accessible and the HCV disease must be controlled by treatment as avoidance techniques, powerful screening programs and globally access to action or treatment [3,11]. On the basis of differences in the heredity changes the HCV strains have been classified into seven genotypes and at least 80 subtypes which vary from each other from 30% to 33% and 20% to 25% respectively. HCV genotypes and subtypes vary in their geographical distribution and they show a high genetic diversity among all genotypes except for genotype 5 and for the recently described genotype 7. HCV genetic variability is not dispersed across all the viral genome that are present. The region that corresponds to the most vital functions such as which of them are involved in translation and replication or which are having more structural domains are the most conserved regions. The 5' are the most conserved region having 90% similarity in sequence identity between the different types of strains. The regions which are located with viral capsid are also a highly conserved region which is having an 81% - 88% of identity in sequence present between the strains of HCV. The sequence that corresponds to the hypervariable regions 1 and 2 of E2 protein is shown less sequence homology having only 50% identity. There are some factors that correspond to their high genetic variability of these viruses having a large number of population size, high replication rates and short generation time (Figure 2) [12,13].

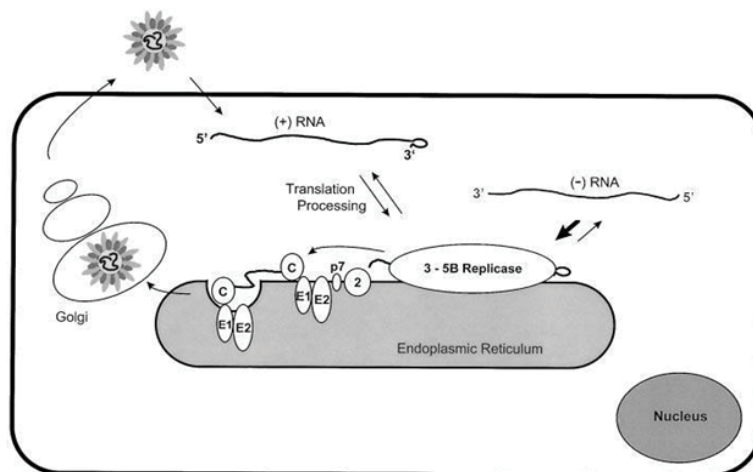


Figure 2 Proposed model of HCV replication cycle (Bartenschlager, et al.) [13]

HCV NS3-4A Serine Protease

The 9.6 kb length of the RNA genome of HCV codes a polyprotein precursor of about 3000 amino acids present which is treated to the cellular as well as viral proteases to the 10 different separate proteins. A non-covalent heterodimer NS3-4A serine protease which contains a catalytic subunit and an activating cofactor. NS4A protein which is responsible for the cleavage at four sites of the polyprotein [13]. The NS3-4A proteases of HCV virus have been considered as one of the target protein which is used for developing novel anti HCV therapy and they are also essential for viral replication. On the other hand, the discovery of selective inhibitors or small molecule against the HCV NS3-4A protease are as oral drug and has been loaded by its substrate-binding groove and the lack of reproducible viral replication models in cell culture [14].

The standard therapy against chronic hepatitis C patients is the combination of pegylated interferon (IFN)- α and daily oral doses of ribavirin [6,10]. Both drugs are indirect antivirals because they do not target a specific HCV protein or RNA element. The therapy is linked with some significant adverse effects including depression, fatigue, and flu-like symptoms which is caused by IFN- α and hemolytic anemia by ribavirin. There is a huge unmet medical need for orally available small molecule and direct anti HCV drugs to provide hepatitis C patients with more effective treatments with fewer side effects.

HCV that are having a long polyprotein precursor of 3000 amino acids which is processed proteolytically upon translation by both cellular and viral proteases to at least 10 individual proteins which includes four structural proteins (C,

E1, E2, and p7) and six nonstructural (NS) proteins (NS2, NS3, NS4A, NS4B, NS5A, and NS5B) (Figure 3). The NS3 is a multi-functional protein having a serine protease domain in its N-terminal one third and a helicase domain in the C-terminal two-thirds. The NS3-4A serine protease is a non-covalent heterodimer complex which is formed by two HCV encoded proteins and the N-terminal serine protease domain of NS3 catalytic subunit and the NS4A cofactor activation subunit. The serine protease is responsible for the proteolytic cleavage at four areas of the HCV polyprotein [15]:

NS3/NS4A

NS4A/NS4B

NS4B/NS5

NS5A/NS5B

The four viral enzymes that encode in the structure of HCV in its nonstructural protein region are NS2-3 auto protease, NS3-4A serine protease, and NS3 helicase and NS3B Serine Protease all of them, they are crucial for HCV replication or infectivity. Among them, NS3-4A serine protease and NS5B RNA dependent RNA polymerase are generally considered to be the attractive targets for the design of new anti-HCV oral drugs [16].

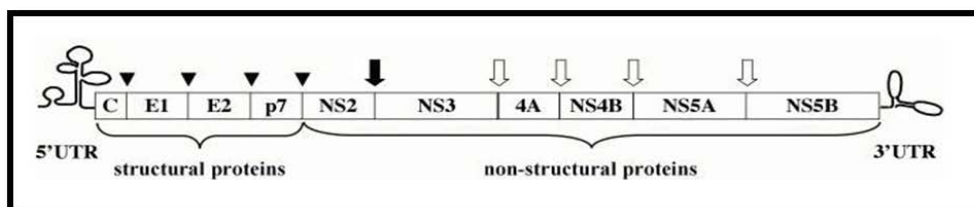


Figure 3 A schematic diagram of the HCV genome

The 5' and 3' untranslated regions (UTR) are shown with secondary structures. The polyprotein encoded by the long ORF is shown as a long box in which individual mature protein products are labeled as core (C), envelope proteins 1 and 2, p7 which are followed by six nonstructural proteins (NS) 2, 3, 4A, 4B, 5A, and 5B. The cleavage sites are marked for cellular signal peptidase HCV NS2-3 auto-protease and NS3-4A serine protease [9].

After the success of the 4th HIV protease inhibitor, it validates that the proteases such as NS3-4A protease could be the great targets for a structure-based drug design approach. However, the substrate-binding groove of the HCV NS3-4A serine protease observed in an X-ray crystal structure and it proposed that the discovery of small molecule and orally available drug candidates would be an extremely difficult job to undertake. Although the lack of a robust and consistent HCV infection cell culture a subgenomic replicon system became the workhouse as the standard assay of antiviral activity of the HCV NS3-4A protease inhibitors (Table 4). In addition to the lack of an HCV infection model in small animals has generally forced to think the scientists to rely on an amalgamation of anti-HCV activity in cell culture and animal pharmacokinetics as substitute indicators of efficacy prior to clinical trials in human [8,17-21].

Table 4 Current HCV NS3/4A Protease based Inhibitors in Pipeline (Kazi, et al.) : NCI: Non-Covalent Inhibitor [9]

Mechanism	Inhibitor Name	Genotype Coverage	DOSE	Company	Status
NCI	Simeprevir (TMC 435)	1 and 2	Once	Janssen	Phase III
NCI	Faldaprevir (BI201335)	1	Once	Boehringer Ingelheim	Phase III
NCI	Danoprevir (RG7227)	1	Twice	Genetech	Phase II
NCI	Vaniprevir (MK-7009)	1	Twice	Merck	Phase II

The HCV NS3 protease is an important source for replication. It is one of the most effective and important attractive targets for developing novel antiviral therapy or drugs. There is a total of two divergent modules of NS3 protease inhibitors that are developed with different mechanism these are Non-covalent classical inhibitors and another one is covalent inhibitors. Resulting in a great success category of covalent inhibitors however the two important clinically tested HCV NS3 protease drug molecules are both keto amides and their scientific name are Boceprevir and Telaprevir. Similarly, we have non-covalent inhibitors where interferon (peg-IFN) is injected with a formulation of Ribavirin

presenting the most effective treatment and acting as an antiviral agent. These non-covalent inhibitors had shown a promising role in safety profiles. Hence preclinical and clinical features have to be described and identified yet [22].

Next-Generation Approach

Next-Generation Sequencing is a high throughput approach that enables a large variety of applications that allow the researchers to study the cell type or organism. The basic mechanism that occurs in NGS technology is the isolation of the DNA. DNA polymerase catalyzes the incorporation of fluorescently deoxyribonucleotide triphosphate into a DNA template during the synthesis of the DNA cycle, on this nucleotide are identified by a fluorophore in every cycle [23]. Whereas the Next Generation Sequencing technology processes millions of fragments in a parallel way as compare to sequencing a single DNA fragment.

Next-generation sequencing introduced the most popular NGS application which is whole-genome sequencing. NGS technology takes research towards the next level towards genomic research. The development of various bioinformatics servers and tools for NGS gives more opportunities toward the research in the genomic field [24]. There are various hands-on experience tools like De novo genome assembly or DNA sequencing to know or understand the genomic variation and most used tools for genomic research (Figure 4). Human genetics has been faced with a new paradigm of research and medical genomics by sequencing technologies since the human project [25].

Next-generation sequencing (NGS) technologies are the technique that is faster and cheaper than Sanger sequencing and DNA sequencing methods a new era for genomic study and molecular biology. If we compared the other Sanger electrophoresis sequencing method which called a first-generation sequencing technology. NGS technologies deliver advanced throughput data with lower cost and enable genome research. NGS technologies have three major enhancements compared to first-generation sequencing [26].

For NGS there is no need to do bacterial cloning and prepare libraries for sequencing in a cell-free system [27].

It processes millions of sequencing reactions at the same time Detection of bases is performed cyclically and in parallel

These improvements allow researchers to process sequencing of the entire genome having low cost and short time. The figure shows the workflows of NGS and conventional methods towards sequencing. The developments of the NGS technology are used for the alignment algorithms to assemble and map the genome from the short reads [27-30]

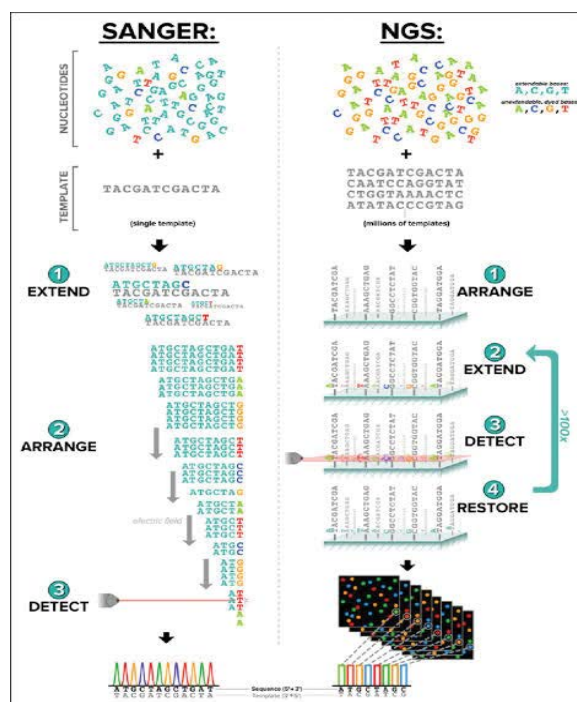


Figure 4 NGS is modified and scaled-up implementation of Sanger sequencing. In both of the methodologies, we have seen that millions of DNA template is sequencing in NGS instead of sequencing a single DNA fragment

Next-Generation Sequencing Mechanism

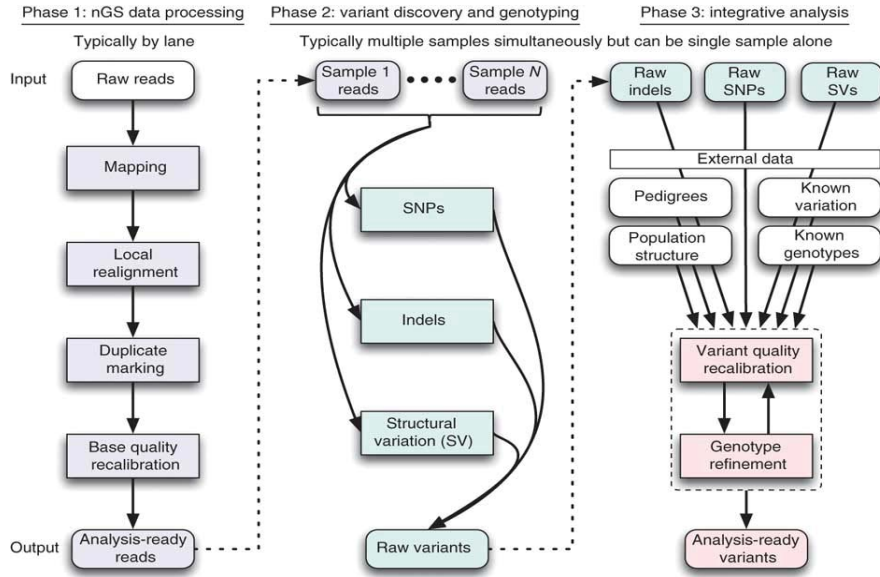


Figure 5 Methods of NGS Data Analysis

The Next-generation sequencing shows four basic steps the figure shows the basic layout of the NGS mechanism Figures 5 and 6.

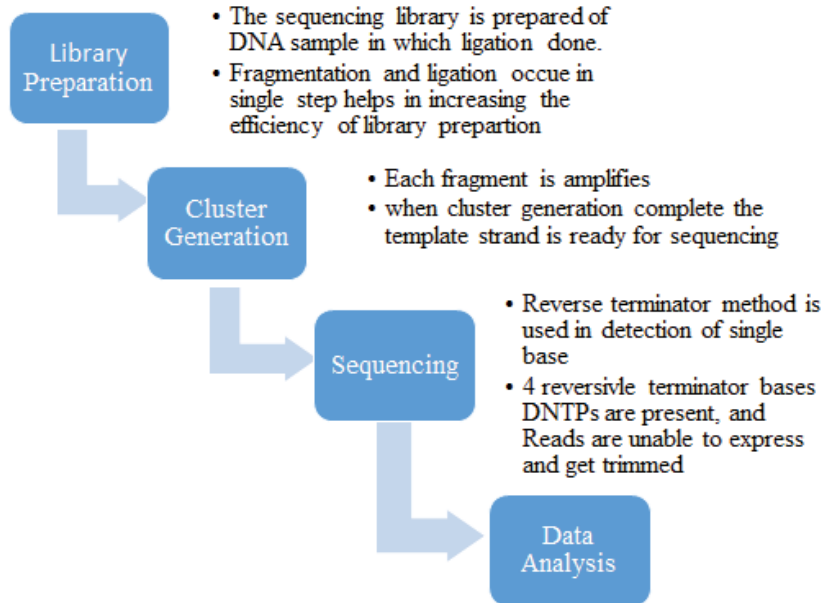


Figure 6 Four major steps that follow during NGS

There are many online servers that are used to analyze the evolution rate as well to find the mutation between the lineages. Practical developments in Next Generation Sequencing offer resources to attain deeper insights into cellular functions [31]. The fewer resources in methodologies become a challenge for the investigation and understanding of RNA sequencing data. For discovering the novel RNA sequences, RNAseq using the NGS technology and for quantifying all transcripts in a cell. Usegalaxy is one of the open resource web servers which enables the key to know and understand the resources and analyzing the genomics and transcriptomics study of viruses [32,33]. From getting high-quality reads and understanding the variance analysis, mutation and annotation use galaxy pipeline is used [34-36].

The online platforms are used to analyze the study of HCV such as NCBI, Consurf, Mutanalyzer, Galaxy, and HCV sequence locator or alignment. The HCV database, The Los Alamos, used to analyze the different strains of HCV. The aim of the study is to understand the evolution rate in different continents, by analyzing the phylogenetic relationships with the other strain of HCV. By using an online server GALAXY for Next Generation Sequencing it takes the NGS to other levels where the GALAXY project helps us to determine and make NGS manipulation, Virus variation and to check the high quality of READS [37-40]. The phylogenetic relation was studied on the HCV Los Alamos database as it utilizes PhyML and offers multiple evolutionary models to infer phylogenetically.

DISCUSSION AND CONCLUSION

The development of science and technologies is growing so fast and they are providing a new way to formulate scientific questions and advance knowledge. Among all of these techniques that are PCR, electron microscope, cell culture, Next Generation Sequencing is one of the technologies that are changing the way and methods to understand the viruses, mostly in the field of genome sequencing, epigenetics, evolution, and transcriptomic analysis. Understanding of these hereditary aspects is a must in all areas of life science research and the introduction of next-generation sequence technology has changed the study of biology. Genomics and molecular biology have continuously been a limitless source of research for global researchers in the field of biology and biotechnology. These two fields have every time, generated a massive amount of data and in order to compile and analyze those, bioinformatics came into action throughout the last few years. Putting into the practice of these bioinformatics tools for the analysis to the sequencing they are fast enough and cut down the cost of laboratory equipment. Next-generation sequencing or high throughput sequencing has assisted a lot to substitute old conventional methods of sequencing and with the help of recent technologies. In genomic studies, a major portion was focused on comparative genomics and genome sequencing. As the human genome project was done during 2003, there was an enormous number of genome sequencing project was observed. This was because all the trouble related to human or other organism is connected to genome composition and variation among these. The HCV infection rate is relatively high with replication range between 10^{10} to 10^{12} virions, and a half-life of 2-4 hours. The HCV RNA mutates rapidly because of the lack of error proofreading by viral RNA polymerase. And these increases in genotype mutation and their subtypes make the research to develop HCV vaccine a challenge.

DECLARATION

Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

- [1] Wasley, Annemarie, and Miriam J. Alter. "Epidemiology of hepatitis C: Geographic differences and temporal trends." *Seminars in Liver Disease*, Vol. 20, No. 1, 2000, pp. 1-16.
- [2] Alter, Miriam J. "The epidemiology of acute and chronic hepatitis C." *Clinics in Liver Disease*, Vol.1, No. 3, 1997, pp. 559-68.
- [3] Soni Abhishek, and Vikas Kaushik. "In silico identification of inhibitors as antagonist for HCV treatment." 2018 International Conference on Bioinformatics and Systems Biology (BSB). *IEEE*, 2018.
- [4] Charlton, Michael. "Hepatitis C infection in liver transplantation." *American Journal of Transplantation*, Vol. 1, No. 3, 2001, pp. 197-203.
- [5] Smith, Donald B., et al. "Expanded classification of Hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource." *Hepatology*, Vol. 59, No. 1, 2014, pp. 318-27.
- [6] Forbi, Joseph C., et al. "Next-generation sequencing reveals frequent opportunities for exposure to Hepatitis C virus in Ghana." *PloS One*, Vol. 10, No. 12, 2015, p. e0145530.
- [7] Seeff, Leonard B., and Jay H. Hoofnagle. "National Institutes of Health consensus development conference: management of Hepatitis C: 2002." *Hepatology*, Vol. 36, No. 1, 2006, pp. S1-S2.

-
- [8] Kim Keun, Young Kim. Centers for disease control and prevention. Recommendations for prevention and control of Hepatitis C virus (HCV) infection and HCV-related chronic disease. *MMWR Recommendations and Reports*, Vol. 47, No. 19, 1998, pp. 1-39.
- [9] Salam, Kazi Abdus, and Nobuyoshi Akimitsu. "Hepatitis C virus NS3 inhibitors: current and future perspectives." *BioMed Research International*, 2013.
- [10] Choo, Qui-Lim, et al. "Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral Hepatitis genome." *Science*, Vol. 244, No. 4902, 1989, pp. 359-62.
- [11] Salmons, M., et al. "First next-generation sequencing full-genome characterization of a Hepatitis C virus genotype 7 divergent subtype." *Clinical Microbiology and Infection*, Vol. 22, No. 11, 2016, p. 947-e1.
- [12] Neumann, Avidan U., et al. "Hepatitis C viral dynamics *in vivo* and the antiviral efficacy of interferon- α therapy." *Science*, Vol. 282, No. 5386, 1998, p. 103-7.
- [13] Bartenschlager, Ralf, and Volker Lohmann. "Replication of Hepatitis C virus." *Journal of General Virology*, Vol. 81, No. 7, 2000, pp. 1631-48.
- [14] Bukh, Jens, Roger H. Miller, and Robert H. Purcell. "Genetic heterogeneity of Hepatitis C virus: quasispecies and genotypes." *Seminars in Liver Disease*. Vol. 15. No. 1, 1995, pp. 41-63.
- [15] Dodd, R. Y., E. P. Notari IV, and S. L. Stramer. "Current prevalence and incidence of infectious disease markers and estimated window-period risk in the American Red Cross blood donor population." *Transfusion*, Vol. 42, No. 8, 2002, pp. 975-79.
- [16] Roth, W. Kurt, et al. "Yield of HCV and HIV-1 NAT after the screening of 3.6 million blood donations in central Europe." *Transfusion*, Vol. 42, No. 7, 2002, pp. 862-68.
- [17] Weiner, Amy J., et al. "Sequence variation in Hepatitis C viral isolates." *Journal of Hepatology*, Vol. 13, 1991, p. S6-S14.
- [18] Schreiber, G. B. "The risk of transfusion-transmitted viral infections. The epidemiology donor study." *N Engl J Med*, Vol. 334, No. 26, 1996, pp.1734-35.
- [19] Alter, Miriam J., et al. "The prevalence of Hepatitis C virus infection in the United States, 1988 through 1994." *New England Journal of Medicine*, Vol. 341, No. 8, 1999, pp. 556-62.
- [20] Baker, Monya. "De novo genome assembly: what every biologist should know." 2012, p. 333.
- [21] Ng, Pauline C., and Ewen F. Kirkness. "Whole-genome sequencing." *Genetic Variation. Humana Press, Totowa, NJ*, 2010, pp. 215-26.
- [22] Wang, Zhong, Mark Gerstein, and Michael Snyder. "RNA-Seq: A revolutionary tool for transcriptomics." *Nature Reviews Genetics*, Vol. 10, No. 1, 2009, p. 57.
- [23] Niu, Wei, et al. "Diverse transcription factor binding features revealed by genome-wide ChIP-seq in *C. elegans*." *Genome Research*, Vol. 21, No. 2, 2011, pp. 245-54.
- [24] European Association For The Study Of The Liver. "EASL International Consensus Conference on hepatitis C: Paris, 26-28, February 1999, consensus statement." *Journal of Hepatology*, Vol. 30, No. 5, 1999, pp. 956-61.
- [25] Mohd Hanafiah, Khayriyyah, et al. "Global epidemiology of Hepatitis C virus infection: New estimates of age-specific antibody to HCV seroprevalence." *Hepatology*, Vol. 57, No. 4, 2013, pp. 1333-42.
- [26] Shi, Stephanie T., et al. "Hepatitis C virus RNA replication occurs on a detergent-resistant membrane that cofractionates with caveolin-2." *Journal of Virology*, Vol. 77, No. 7, 2003, pp. 4160-68.
- [27] Nkrumah, B., et al. "Hepatitis B and C viral infections among blood donors from rural Ghana." *Ghana Medical Journal*, Vol. 45, No. 3, 2011.
- [28] Galagan, James E., et al. "The *Mycobacterium tuberculosis* regulatory network and hypoxia." *Nature*, Vol. 499, No. 7457, 2013, p. 178.

-
- [29] Schmutz, Jeremy, et al. "Quality assessment of the human genome sequence." *Nature*, Vol. 429, No. 6990, 2004, p. 365.
- [30] Echeverría, Natalia, et al. "Hepatitis C virus genetic variability and evolution." *World Journal of Hepatology*, Vol. 7, No. 6, 2015, p. 831.
- [31] Bukh, Jens, Robert H. Purcell, and Roger H. Miller. "Sequence analysis of the 5' noncoding region of hepatitis C virus." *Proceedings of the National Academy of Sciences*, Vol. 89, No. 11, 1992, pp. 4942-46.
- [32] Argentini, Claudio, et al. "HCV genetic variability: From quasispecies evolution to genotype classification." 2009, pp. 359-73.
- [33] Lavanchy, D. "Evolving epidemiology of Hepatitis C virus." *Clinical Microbiology and Infection*, Vol. 17, No. 2, 2011, pp. 107-15.
- [34] Sandres-Saune, K., et al. "Determining the Hepatitis C genotype by analyzing the sequence of the NS5b region." *Journal of Virological Methods*, Vol. 109, 2, 2003, pp.187-193.
- [35] Knipe, D. M., and P. M. Howley. "Flaviviridae: The virus and their replication." *Fields virology, 4th ed. Lippincott-Raven, Philadelphia, PA*, 2001, pp. 991-1029.
- [36] Qiu, Ping, et al. "HCV genotyping from NGS short reads and its application in genotype detection from HCV mixed infected plasma." *PloS One*, Vol. 10, No. 4, 2015, p. e0122082.
- [37] Zanini, Barbara, and Alberto Lanzini. "Antiviral treatment for chronic Hepatitis C in illicit drug users: A systematic review." *Antiviral Therapy*, Vol. 14, No. 4, 2009, p. 467-79.
- [38] Sy, Theodore, and M. Mazen Jamal. "Epidemiology of Hepatitis C virus (HCV) infection." *International Journal of Medical Sciences*, Vol. 3, No. 2, 2006, p. 41.
- [39] Besaratinia, Ahmad, et al. "A high-throughput next-generation sequencing-based method for detecting the mutational fingerprint of carcinogens." *Nucleic Acids Research*, Vol. 40, No. 15, 2012, p. e116.
- [40] Muzzey, Dale, Eric A. Evans, and Caroline Lieber. "Understanding the basics of NGS: From mechanism to variant calling." *Current Genetic Medicine Reports*, Vol. 3, No. 4, 2015, pp. 158-65.