

# Exploring the complexity of RNA virus infections using viral genomics

Chaturaka Rodrigo<sup>1</sup>

*Journal of the Ceylon College of Physicians*, 2021, **52**, 71-77

## Abstract

RNA viruses such as influenza, dengue, SARS, Ebola and yellow fever are a formidable threat to human health and global economy. They have small genomes, but due to the lack of proof-reading during RNA-to-RNA replication, these genomes harbour many mutations. Thus, in a single infected host, many different genomic variants exist. Until recently, it was not possible to sequence these within-host variants reliably due to technical limitations of sequencing platforms. Instead, most analyses were restricted to a representative “consensus” viral sequence per host, which is an oversimplification of the true diversity of these infections. While consensus level genome analyses have helped in understanding disease pathogenesis, designing better treatments or vaccines, contact tracing and predicting infection outcomes, using within-host variants in analyses could potentially achieve more. The work presented here describes the value of viral genomics in understanding the complexity of RNA virus infections and introduces cutting-edge technological advances we have made in identifying within-host viral variants in an infected host. This work used hepatitis C virus (a chronic infection) as a model and the optimised methods were subsequently adapted to study dengue virus (an acute infection). Chronic and acute RNA virus infections each pose unique challenges in viral genomics analyses and overcoming these challenges for both types of viruses ensures that these methods can be applied to study many other RNA viruses.

## Introduction

Pathogenic RNA viruses such as dengue, yellow fever, influenza, Ebola, hepatitis C, SARS-1 and more recently SARS-nCoV-2, have posed significant threats to human health and global economy through epidemics and pandemics.<sup>1-3</sup> Arguably, as a group, these are the most significant contemporary threat to human health in the domain of infectious diseases.<sup>4</sup> Viral genomics are an important tool to track these infections across hosts, understand their pathogenesis, explore pathogen interactions with host immunity, and predict outcomes of infection (with or without treatment).<sup>1, 2, 5, 6</sup> This oration will highlight my work in collaboration with others in the field of viral genomics over the past seven years.

RNA viruses have small genomes. For example, SARS-nCoV-2 genome which is only 30kbs in length, is one of the longest known pathogenic RNA virus genomes.<sup>7</sup> However, given the lack of proofreading during RNA-to-RNA genome replication, the mutagenesis within these genomes is far greater than that in the more “stable” but longer DNA genomes.<sup>1</sup> In fact, mutations in RNA viruses are said to occur in “real-time”, with significant changes to the genetic code being observed within days, weeks, or months. In contrast, in DNA virus genomes where mutations happen 100-10000 times slower, significant changes may not be visible for a much longer time. Even in a short-lived RNA virus infection like dengue, there is no “single virus genome” within an infected host. Instead,

<sup>1</sup>Senior Lecturer in Pathology, School of Medical Sciences, Faculty of Medicine and Health, University of New South Wales, Sydney, Australia.

Correspondence: E-mail: [c.rodrido@unsw.edu.au](mailto:c.rodrido@unsw.edu.au)

 <https://orcid.org/0000-0003-2189-9177>

Received 20 October 2021, accepted 30 November 2021.



This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

there are multiple genomic variants (quasi-species or within host variants)<sup>8-10</sup> which for logistical and scientific reasons are often not accounted for in many viral genomics studies. Instead, a representative sequence per host called the “consensus sequence” is constructed to represent all variants in an infected host but this is an oversimplification the true diversity of the infection. The main reason for using consensus sequences is the technical difficulty in amplifying entire viral genomes (of variants) as intact molecules and sequencing them without fragmenting. With the introduction of next generation sequencing (NGS),<sup>11</sup> it became possible to identify low frequency single nucleotide polymorphisms (SNPs) in RNA or DNA genomes but the popular technologies such as paired-end short read sequencing (Illumina platform), fragments genomes during sequencing, and it is impossible to accurately reconstruct original variants lengthwise from this sequencing output. This oration will focus on our efforts undertaken between 2014- 2021 to develop assays to amplify intact whole genomes, sequence them without fragmenting and using the output to identify within-host variants. Furthermore, the impact of large-scale viral genome sequencing projects in advancing the knowledge in disease epidemiology, pathogenesis and outcome prediction will be discussed. This work used hepatitis C virus (HCV) as a model which is an average length (9-10kb) RNA virus genome, but the methods are transferable to study other viruses as demonstrated for the dengue virus, at the end. For clarity, each project will be presented with methods and results combined, starting from analyses that used consensus sequences only, followed by those which also analysed the nucleotide polymorphisms (SNPs) and finally, the cutting-edge projects that identified full-genome length within-host variants. All the studies mentioned in this section have been conducted under appropriate ethical approvals from University of New South Wales, Australia (HC14201, HC190074, HC180015).

\*Within-host variants mentioned in this paper are the near-full length viral variant genomes within an infected host (sometimes referred to as haplotypes). Individual mutations at a given genome position that differ from the consensus are referred to as single nucleotide polymorphisms (SNPs). SNPs can be accurately and easily characterised by most next generation sequencing platforms (e.g., Illumina). Some papers refer to individual SNPs as “variants”, and this must not be confused with the variants mentioned in this paper. Characterising true within-host variants (different SNP combinations on the same viral genome) is more challenging.

### **Generating full length HCV amplicons as intact molecules**

A pre-requisite to identifying within host Hepatitis C variants is to develop an assay to amplify the entire HCV genome as a single molecule. We achieved this aim in 2016,<sup>12</sup> marking a significant departure from the standard assays at that time which amplified viral genomes as overlapping fragments before concatenating these to create a consensus sequence. Our assay used long range Taq polymerase enzymes in a polymerase chain reaction (PCR) after reverse transcription to generate a near full-length, double stranded amplicon of the HCV genome. This new assay was successfully tested with 122 samples belonging to 6 major HCV genotypes (and the newly discovered genotype 8 in 2017-18).<sup>12</sup> The lower cut-off viral load for successful full-length genome extraction was 14,850 IU/ml (with a 90% success rate above this cut-off). This method was also high throughput and cost effective (AUD 100 per sample in 2016) compared to contemporary methods of HCV genome extraction and sequencing. However, we were unable to sequence these genomes as whole molecules because the paired-end short read NGS technology (Illumina) popular at that time processed its output (reads) as nucleotide fragments of 300-600nts in length. Using these fragmented read outputs, we were only able to generate a consensus sequence per sample (and the frequencies of each individual SNPs). This problem was finally solved three years later (see below).

### **Cross-continental phylogenetics of HCV evolution in acute infection**

Armed with a cheaper, cost effective and a high throughput sequencing pipeline, we took on an ambitious project to generate an HCV sequence database from samples collected during the acute phase of infection (first 6 months of infection). From a genomics perspective, early acute infection is very interesting because during this time a series of bottlenecks in viral genomic diversity occur either leading to chronic infection or spontaneous clearance (observed in approximately 25% of patients). Therefore, establishing an acute infection sequence repository was required to study these changes in detail. However, early HCV infection is asymptomatic and such clinical samples are vanishingly rare. Thus, to achieve a significant sample size, an international collaboration was necessary. This requirement was fulfilled by the InC3 consortium which is a collection of nine prospective cohort studies that recruited intravenous drug users in the Netherlands, Australia, United States and Canada to identify incident HCV infections between 1990-2012.<sup>13</sup> Collectively, these cohorts had 369 incident HCV cases and we managed to sequence

213 genomes from these, establishing the world's largest acute infection HCV sequence database: InC3 viral sequence repository (InC3-VSR).<sup>13</sup> Using consensus sequences from this repository, we proved that using near-full-length genomes in phylogenetics (inferring related infections via genetic sequence similarity) is superior to using smaller segments of the genome, in terms of improving the resolution of the analysis. Phylogenetic analyses from this database also demonstrated that sequences from North America and Australia clustered separately (implying that patterns of mutations observed have a geographical bias), suggesting that the virus may be evolving in separate micro-epidemics in geographically "isolated" communities.<sup>13</sup> This was a concern since it could translate to a geographical bias in mutations which confers drug resistance.

### **Inferring history of HCV community spread from viral sequences**

If a rapidly mutating virus is spreading in a community, as the number of infected people increase overtime, the genomic diversity of the viruses isolated from patients should also increase (combined effect of within and between-host evolution).<sup>14,15</sup> Conversely, viral genomic diversity overtime can be a surrogate measure of the number of infected people in a community, when the latter is unknown. This feature can be used to estimate fluctuations of the infected population size overtime (into the past), by observing temporal changes in viral genomic diversity (a phylogenetic tree with a time axis).<sup>16</sup> The tips of such a tree represent actual sequences while the nodes of the tree represent their ancestors (hypothetical). The size of a simulated infected population (and its changes overtime) could be inferred by the branching pattern of this tree on the assumption that a shrinking phylogenetic diversity (less branching) indicates a lower number of infections. We used consensus sequences from the InC3-VSR to generate phylogenetic trees with a time axis to understand how the HCV infected population size in North America and Australia fluctuated overtime.<sup>17</sup> These analyses were done using a software suite for evolutionary analyses based on Bayesian statistics (Bayesian evolutionary analysis and sampling of trees – BEAST version 1.8)<sup>16</sup>. We demonstrated that the origin of HCV subtype 1a infections was earlier (around 1920) than that of subtype 3a (around 1950) in both continents and that epidemics of both subtypes saw an exponential increase of infections between 1955-1975 before slowing and stabilising in the 1990s.<sup>17</sup> This model was validated by epidemiological estimates from other investigators.<sup>18</sup> This study demonstrates the value of sequence based phylogenetics in interrogating historical changes that

led to fluctuations in rates of infection. Lessons learnt from such exercises are useful to develop better surveillance to prevent epidemics.

### **Understanding impact of host immunity in acute infection through viral genomics**

Viral sequencing data, especially that of the coding regions for T and B cell epitopes, provide an ideal opportunity to explore or predict the interaction of the virus and the host immune system. Broadly neutralising antibodies (antibodies that can neutralise multiple genotypes of HCV) are thought to play an important role in spontaneous clearance of infection observed in some individuals. We downloaded and analysed 1749 HCV consensus genomes isolated during the chronic phase of infection from publicly available databases (all available sequences at the time of analysis) to explore mutations in the viral genomic regions that code for binding domains of broadly neutralising antibodies.<sup>19</sup> We then compared these with acute infection sequences from the InC3-VSR to see if some mutations conferring resistance to broadly neutralising antibodies (thus making virus neutralisation less effective) are specific to the acute stage of infection. We found that contact residues for all known broadly neutralising antibodies were restricted to three linear regions of the HCV Envelope 2 protein.<sup>19</sup> Experimentally proven resistant mutations to antibody binding (identified *in vitro*) were rare in naturally circulating sequences isolated from patients (*in vivo*). For example, only 10 out of 29 known resistant mutations had a frequency of occurrence greater than 5% among circulating sequences. Two sites (positions 610 and 655 in E2 protein) within broadly neutralising antibody epitopes had significant differences in the type of amino acid residue seen in between acute and chronic infection sequences.<sup>19</sup>

### **Naturally occurring drug resistant mutations in treatment naïve patients**

All of the above-mentioned projects were based on consensus sequences. However, as all InC3 sequences were subjected to next generation sequencing, even rare mutations seen in some within-host variants could be quantified. Any mutation in a sequenced read that is different to the consensus sequence base at the same position is called single nucleotide polymorphism (SNP). This and the following project was on SNP quantification.

Since 2012, availability of highly successful directly acting antiviral drugs (DAAs) have revolutionized HCV therapy, offering >90% cure rates with minimal side effects.<sup>20</sup> However, in a minority, first line DAAs

fail due to drug resistance. These resistant SNPs (Resistance associated substitutions -RASs) are coded in NS3, NS5A and NS5B regions of the HCV genome. Given that InC3-VSR had a number of acute infection sequences from four different countries isolated between 1990-2012, and as each of these genomes were deep sequenced to an average depth of 17,000 reads per nucleotide position, naturally occurring RASs with an abundance as low as 0.1% could be reliably identified.<sup>21</sup> Furthermore, in one cohort patients were longitudinally sequenced giving the opportunity to see if RASs spontaneously appear and disappear overtime, even when not exposed to DAAs.<sup>22</sup> We found that in InC3-VRS, naturally occurring RASs against more than one DAA (or their combinations) were rare. Therefore, had these individuals been treated, even in the unlikely event of resistance to a first line drug, a second-line drug would have been successful. Interestingly some naturally occurring RASs were significantly more frequent in some genotypes while two important RASs (NS3 Q80K in genotype 1a and NS5B N142T in genotype 3a) had a geographical bias in occurrence also.<sup>21</sup> In the longitudinal sample analysis, we demonstrated some RASs may disappear overtime when they occur on T cell epitopes, due to the host immune selection pressure.<sup>22</sup>

#### **Determinants of genetic variability in HCV**

To understand the factors driving the high mutation rate of HCV and to see if the mutation rates can predict infection outcomes such as clearance, we computed the “mutability” of HCV genome using “Shannon entropy” which is a measure of uncertainty quantified as a value between 0 and 1 with zero indicating immutable, highly conserved positions. We compared Shannon Entropy values across the full genome and different sections of the genome and correlated these with host genetic polymorphisms, country of origin, host gender and infection outcome (spontaneous clearance or chronic infection) to see if viral genetic diversity is associated with any of these factors.<sup>23</sup> We found that most of the genomic mutations in the acute phase of infections were non-synonymous mutations (leads to a change in the coded amino acid), and that the Envelope (important for viral entry to hepatocytes) and NS5B regions (important for viral replication) of the genome had the highest mutability.<sup>23</sup> Viral genotype 1a (vs 3a) and host IFNL3 allele CC polymorphism at position rs12979860 (vs non-CC genotypes) were associated with a higher genomic variability of the virus thus, demonstrating both host and virus-related factors may drive mutations in the HCV genome. Host IFNL3 gene is important for innate immunity and its polymorphisms have been linked to spontaneous clearance of HCV by other authors.<sup>24</sup>

#### **Limitations of short-read sequencing**

With the paired-end short read NGS platform (Illumina) which was used to generate the InC3-VSR, it is possible to generate an accurate consensus sequence per sample, and identify single nucleotide polymorphisms (SNPs) accurately, but how these SNPs are combined on a genome to create a unique variant could not be visualised as this platform generates short reads of 300-600 nucleotides in length (HCV genome is 9.5kb in length). Probability based bioinformatic algorithms can “reconstruct” a longer genomic segment from these short reads (haplotype reconstruction) based on shared mutations on overlapping parts of the short reads.<sup>25-27</sup> However, these algorithms have poor agreement with each other and in absence of a gold standard, it is not possible to validate if the reconstructed haplotypes are real.<sup>28</sup> In addition, the errors of these algorithms increase with the reconstructed length of the genome and hence it is not possible to reconstruct full-genome length haplotypes. We observed these problems firsthand in a study where we attempted to characterise HCV founder variants using different haplotype reconstruction algorithms from InC3-VSR data.<sup>29</sup> Founders are distinct within-host variants of viruses that can transmit an infection from one host to another. The existence of founder variants has been demonstrated in several RNA virus infections including HCV, dengue and influenza. We used different haplotype reconstruction algorithms focussing on three distinct regions of the HCV genome (each approximately 2000nts in length).<sup>29</sup> Previous studies that defined transmitted founders used one haplotype reconstruction algorithm and focussed on the Envelope region of the HCV genome only, and we wanted to see if this method is accurate once results are cross validated with different algorithms and across other regions of the genome. After assessing 190 samples, 54 very early acute infection samples were identified as potentially having founder variants, but after cross validating across the two algorithms and three different regions of the genome, only 14 single transmitted founders could be identified with certainty.<sup>29</sup> While we identified a higher number of precious founder sequences with greater certainty than any of the previous studies by others, this study also demonstrated the uncertainty and unreliability of haplotype reconstruction algorithms as they do not produce concordant results on most occasions.

#### **Moving beyond haplotype reconstruction to identify within host variants**

As we were exploring ways to overcome the limitations of next generation sequencing to identify

within host variants reliably, long read sequencing technologies, sometimes referred to as “Third generation sequencing” (e.g., Single Molecule Real-time sequencing technology – SMRT, Oxford Nanopore Technology – ONT) entered the space of commercial sequencing. These new technologies generate individual reads of up to 100kbs in length, far greater than the length of an RNA virus genome. Thus, for the first time there was a chance to sequence entire within-host viral variants, bypassing haplotype reconstruction. In addition, ONT had the added advantage of using a hand-held portable sequencer (MinION), enabling remote sequencing without a sophisticated laboratory, markedly cutting down the cost of sequencing.<sup>7,30-33</sup> However, compared to Illumina (short read) platform, these newer technologies had a high error rate, and when sequencing RNA viruses the difficulty in differentiating a true mutation from a sequencing error posed a significant problem. In an extensive body of work spanning over three years, we demonstrated the suitability of utilising ONT for RNA virus sequencing when certain conditions were met (using sequencing depth to balance the sequencing error).<sup>34</sup> In other words, we showed that for the same sample, the consensus sequences generated by ONT and Illumina platforms were similar if the ONT consensus was built from at least 300 individual reads or more.<sup>34</sup> We also developed a novel bioinformatics tool named Nano-Q (<https://github.com/PrestonLeung/Nano-Q>) which uses a hierarchical clustering algorithm, to differentiate within host variants from ONT sequencing while adjusting for sequencing errors.<sup>34</sup> To test our workflow, we mixed plasmid clones with known HCV sequence inserts in known proportions to recreate a “variant mix” and then applied the new ONT based workflow to see if it can identify constituent plasmid HCV sequences and their proportions correctly. We demonstrated that our workflow reproduced all variants in the mix even when one constituent had an abundance as low as 0.1%. Overall, ONT sequencing was cheaper compared to Illumina sequencing with the average cost per sample being around AUD 43 (vs AUD 100) in 2020.

### **Beyond hepatitis C; Studying other RNA viruses**

Hepatitis C is a highly mutable virus which accumulates many mutations overtime given the chronic nature of infection. Genomes of short-lived infections such as dengue or SARS-nCoV-2 have fewer mutations but pose new challenges when sequenced with an error prone platform such as ONT as sequencing errors heavily influence relationships seen in a phylogenetic tree in a low mutation environment (risk of inflating genetic diversity across variants by sequencing errors). We have worked on the dengue virus on the assumption that if our optimised ONT

workflow can be applied to both a chronic (HCV) and an acute infection (dengue), then it can be adapted to study many other RNA viruses. Thus, we developed a novel assay to amplify full genome dengue sequences as single molecules, developed an ONT based workflow for cost effective high throughput dengue virus sequencing and adapted the Nano-Q tool to differentiate within host variants of dengue virus.<sup>35</sup> Finally, we established pre-conditions to be followed when using sequences with less mutations (e.g., dengue) generated from an error prone platform (ONT) for phylogenetics. Currently this pipeline is being used to generate a large dengue sequencing dataset from an ongoing cohort study in Sri Lanka; The Colombo Dengue study.<sup>36,37</sup>

### **Discussion**

The extensive body work of presented above made significant advances in understanding the pathogenesis of HCV infection from a viral genomics perspective, while making significant technological gains to “re-set” the standards in this field. The full genome amplification method developed by us has been used by other teams to design similar assays for other viruses such as enteroviruses and SARS-nCoV-2.<sup>7,38</sup> The findings from above mentioned projects had a significant knowledge impact on understanding HCV pathogenesis, predicting outcomes and designing elimination strategies. For example, the finding of a strong geographical bias in phylogenetics of HCV evolution and that such variations were mainly located in envelope and NS5B regions (rich in T cell epitopes), questioned if a universal vaccine would have the same effect across different geographic regions. The historical reconstructions of HCV epidemics in North America and Australia demonstrated how social circumstances shaped these epidemics and the value of such modelling to track the ongoing epidemics in the DAA era, to see if the infected population is declining with successful community-based treatment. Characterising epitopes and their differences in acute and chronic infections is an important step to further explore the mechanisms of spontaneous clearance (which is still unknown) as well as to design successful vaccines against HCV. Examination of naturally occurring RASs against DAAs across multiple genotypes and countries revealed that RASs conferring resistance to multiple drug regimens are rare and even if treatment with a first-line option fails, a second line option should be effective – a finding that resonates with real-world experience in HCV treatment. Exploration into mutability of HCV genome showed that such variation is shaped by both viral genotype as well as human allelic genotypes (thus an

interaction of viral and host genomes). This is not an incidental finding as the same human IFNL3 allele has been linked to spontaneous clearance of HCV by other authors.<sup>24</sup>

However, further progression in this field was hindered due to the inability to reliably sequence within host viral variants.<sup>25,26</sup> Including within host variants (instead of a single consensus) into genomic analyses will add another layer of complexity but it will also cause a paradigm shift in the field by allowing us to observe origin and extinction of biologically important variants amidst the selection pressures imposed by host immunity and treatment. This will enable better drug and vaccine design, prediction of treatment outcome, monitoring of emerging drug resistance and pre-emptive changes in treatment before the resistant variant becomes dominant. We have initiated this process by overcoming the hurdle of intact full genome amplification and sequencing, paving the way for reliable differentiation of within host variants without haplotype reconstruction.

The importance of viral phylogenetics has come to the fore more than ever with the SARS-nCoV-2 pandemic. In Australia, phylogenetics is an essential part of its successful contact tracing that identified emerging clusters in real-time.<sup>7</sup> Even in chronic infections such as HIV and HCV, real-time detection of emerging transmission clusters for rapid public health response is gaining traction as a prevention strategy in Australia. For other endemic arboviral infections such as dengue, community surveillance of circulating variants is useful in predicting changes of disease phenotype. However, these aspects are largely unexplored in low-middle- income countries due to the difficulties in establishing high-end sequencing facilities. With emergence of cheaper and portable technologies such as ONT, this situation is changing and scientists as well as clinicians should be aware of the advantages and translational value of viral genomic analyses.

### Acknowledgements

I gratefully acknowledge my collaborators and co-authors from the InC3 consortium (University of New South Wales, Australia, University of Melbourne, Australia, University of New Mexico, USA, Harvard University, USA, Johns Hopkins Medical Institutions, USA, University of Montreal, Canada, Public Health Service and Academic Medical Centre of Amsterdam, The Netherlands) for their contributions for the HCV studies and collaborators from University of Colombo, Sri Lanka, Agency for Science, Technology and Research, Singapore, and National University of

Singapore for contributions to dengue studies. I am grateful for the guidance and supervision of Professors Senaka Rajapakse and Deepika Fernando (for dengue studies) and Andrew Lloyd, Rowena Bull and Fabio Luciani (for HCV studies). The work presented here was also contributed by my PhD students at UNSW (Arun Abayasingham, Nasir Riaz, Braulio Valencia, Hossain Sazzad, Thiruni Adikari), Ms. Chathurani Sigera (PhD student, University of Colombo) and Dr. Preston Leung (Bioinformatician, University of Copenhagen, Denmark). My research is funded by grants from the University of New South Wales and the National Health and Medical Research Council of the Australian Government (No. 1173666). Finally, none of this work would have been possible without the generosity and goodwill of patients who donated their samples and data for these projects.

### References

1. Rodrigo C, Luciani F. Dynamic interactions between RNA viruses and human hosts unravelled by a decade of next generation sequencing. *Biochim Biophys Acta Gen Subj*. 2019; **1863**(2): 511-9.
2. Stanaway JD, Shepard DS, Undurraga EA, Halasa YA, Coffeng LE, Brady OJ, et al. The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. *Lancet Infect Dis*. 2016; **16**(6): 712-23.
3. Tang JW, Lam TT, Zaraket H, et al. Global epidemiology of non-influenza RNA respiratory viruses: data gaps and a growing need for surveillance. *Lancet Infect Dis*. 2017; **17**(10): e320-e6.
4. Woolhouse MEJ, Brierley L. Epidemiological characteristics of human-infective RNA viruses. *Scientific Data*. 2018 02/20/online; **5**: 180017.
5. Liu SQ, Deng CL, Yuan ZM, Rayner S, Zhang B. Identifying the pattern of molecular evolution for Zaire ebolavirus in the 2014 outbreak in West Africa. *Infect Genet Evol*. 2015; **32**: 51-9.
6. Bhatt S, Holmes EC, Pybus OG. The genomic rate of molecular adaptation of the human influenza A virus. *Mol Biol Evol*. 2011; **28**(9): 2443-51.
7. Bull RA, Adikari T, Hammond JM, et al. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *bioRxiv*. 2020: 2020.08.04.236893.
8. Eigen M, McCaskill J. Molecular Quasi-species. *J Phys Chem*. 1988; **92**: 6881-91.
9. Capobianchi MR, Giombini E, Rozera G. Next-generation sequencing technology in clinical virology. *Clin Microbiol Infect*. 2013; **19**(1): 15-22.
10. Bull RA, Luciani F, McElroy K, et al. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. *PLoS Pathogens*. 2011; **7**(9): e1002243.

11. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016; **17**(6):333-51.
12. Bull RA, Eltahla AA, Rodrigo C, Koekkoek SM, Walker M, Pirozyan MR, et al. A method for near full-length amplification and sequencing for six hepatitis C virus genotypes. *BMC genomics.* 2016; **17**: 247.
13. Rodrigo C, Eltahla AA, Bull RA, Luciani F, Grebely J, Dore GJ, et al. Phylogenetic analysis of full-length, early infection, hepatitis C virus genomes among people with intravenous drug use: the InC(3) Study. *J Viral Hep.* 2017; **24**(1): 43-52.
14. Drummond AJ, Rambaut A. Bayesian evolutionary analysis by sampling trees. In: Lemey P SM, Vandamme A., editor. *The Phylogenetic Handbook: a Practical Approach to Phylogenetic Analysis and Hypothesis Testing.* 2 ed. Cambridge: Cambridge University Press; 2009. p. 564-91.
15. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol.* 2005; **22**(5): 1185-92.
16. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 2007; **7**: 214.
17. Rodrigo C, Eltahla AA, Bull RA, Grebely J, Dore GJ, Applegate T, et al. Historical Trends in the Hepatitis C Virus Epidemics in North America and Australia. *J Infect Dis.* 2016; **214**(9): 1383-9.
18. Center for Disease Analysis. *Observatory POLARIS.* 2015 [cited 2016 March 3]; Available from: <http://centerforda.com/polaris/dash.htm>
19. Rodrigo C, Walker MR, Leung P, et al. Limited naturally occurring escape in broadly neutralizing antibody epitopes in hepatitis C glycoprotein E2 and constrained sequence usage in acute infection. *Infection, genetics and evolution: Infect Genet Evol.* 2017; **49**: 88-96.
20. Dore GJ. The changing therapeutic landscape for hepatitis C. *Med J Aust.* 2012; **196**(10): 629-32.
21. Eltahla AA, Rodrigo C, Betz-Stablein B, et al. Analysis of resistance-associated substitutions in acute hepatitis C virus infection by deep sequencing across six genotypes and three continents. *J Viral Hep.* 2017; **24**(1): 37-42.
22. Eltahla AA, Leung P, Pirozyan MR, Rodrigo C, et al. Dynamic evolution of hepatitis C virus resistance-associated substitutions in the absence of antiviral treatment. *Scientific Reports* 2017; **7**: 41719.
23. Rodrigo C, Leung P, Lloyd AR, et al. Genomic variability of within-host hepatitis C variants in acute infection. *J Viral Hep.* 2019; **26**(4): 476-84.
24. Grebely J, Page K, Sacks-Davis R, et al. The effects of female sex, viral genotype, and IL28B genotype on spontaneous clearance of acute hepatitis C virus infection. *Hepatology* 2014; **59**(1): 109-20.
25. Beerenwinkel N, Gunthard HF, Roth V, Metzner KJ. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. *Front Microbiol.* 2012; **3**: 329.
26. Beerenwinkel N, Zagordi O. Ultra-deep sequencing for the analysis of viral populations. *Curr Opin Virol.* 2011; **1**(5): 413-8.
27. Topfer A, Marschall T, Bull RA, Luciani F, Schonhuth A, Beerenwinkel N. Viral quasispecies assembly via maximal clique enumeration. *PLoS Comput Biol.* 2014; **10**(3): e1003515.
28. Pandit A, de Boer RJ. Reliable reconstruction of HIV-1 whole genome haplotypes reveals clonal interference and genetic hitchhiking among immune escape variants. *Retrovirology* 2014; **11**: 56.
29. Abayasingam A, Leung P, Eltahla A, et al. Genomic characterization of hepatitis C virus transmitted founder variants with deep sequencing. *Infect. Genet. Evol.* 2019; **71**: 36-41.
30. Castro-Wallace SL, Chiu CY, John KK, Stahl SE, Rubins KH, McIntyre ABR, et al. Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *Scientific Reports* 2017; **7**(1): 18022.
31. Hoenen T. Sequencing of Ebola Virus Genomes Using Nanopore Technology. *Bio Protoc.* 2016; **6**(21).
32. Ip CL, Loose M, Tyson JR, de Cesare M, Brown BL, Jain M, et al. MiniION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000 Research* 2015; **4**: 1075.
33. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 2016 2016/11/25; **17**(1): 239.
34. Riaz N, Leung P, Barton K, Smith MA, Carswell S, Bull R, et al. Adaptation of Oxford Nanopore technology for hepatitis C whole genome sequencing and identification of within-host viral variants. *BMC Genomics.* 2021; **22**(1): 148.
35. Adikari TN, Riaz N, Sigera C, Leung P, Valencia BM, Barton K, et al. Single molecule, near full-length genome sequencing of dengue virus. *Scientific Reports.* 2020; **10**(1): 18196.
36. Sigera PC, Amarasekara R, Rodrigo C, et al. Risk prediction for severe disease and better diagnostic accuracy in early dengue infection; the Colombo dengue study. *BMC Infect Dis.* 2019; **19**(1): 680.
37. Sigera PC, Rajapakse S, Weeratunga P, et al. Dengue and post-infection fatigue: findings from a prospective cohort-the Colombo Dengue Study. *Trans R Soc* 2020 Oct 24.
38. Isaacs SR, Kim KW, Cheng JX, et al. Amplification and next generation sequencing of near full-length human enteroviruses for identification and characterisation from clinical samples. *Scientific Reports.* 2018 2018/08/08; **8**(1): 11889.