medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

TITLE PAGE

A short and informative title:

A comprehensive five-year study of Hepatitis C Virus infection and its transmission dynamics in multi-transfused β-globin defective patients in the state of West Bengal, India

The full name(s) of the author(s):

Supradip Dutta[1*], Aritra Biswas[1*], Promisree Choudhury[1*], Prosanto Chowdhury[2], Maitreyee Bhattacharyya[3], Sharmistha Chakraborty[4], Shanta Dutta[1], Provash C Sadhukhan[1]

[*] ***Authors contributed equally***

The affiliation(s) and address(es) of the author(s):

[1] Indian Council of Medical Research-National Institute of Cholera and Enteric Diseases
P-33, Scheme XM, CIT Road, Beliaghata, Kolkata-700010, West Bengal, India.

[2] Peerless Hospital & B. K. Roy Research Centre
360 Panchasayar**,** Kolkata – 700094, West Bengal, India.

[3] Medical College Kolkata
88, College St, College Square, Kolkata- 700073, West Bengal, India.

[4] Burdwan Medical College and Hospital
Baburbag, Rajbati, Burdwan – 713104, West Bengal, India.

The e-mail address, and telephone number(s) of the corresponding author:

Provash Chandra Sadhukhan,
Email ID: provash2000@gmail.com
Tel: 91-9830546338

Author contributions:

Conceptualization: Provash C. Sadhukhan; Methodology: Aritra Biswas, Supradip Dutta; Formal analysis and investigation: Aritra Biswas, Promisree Choudhury, Supradip Dutta, Provash C. Sadhukhan; Writing - original draft preparation: Promisree Choudhury; Writing - review and editing: Promisree Choudhury, Aritra Biswas, Supradip Dutta, Provash C. Sadhukhan; Funding acquisition: Provash C. Sadhukhan; Resources: Prosanto Chowdhury, Maitreyee Bhattacharyya, Sharmistha Chakraborty, Shanta Dutta, Provash C Sadhukhan; Supervision: Provash C. Sadhukhan.

Declaration of Interests: None

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Highlights:

- HCV prevalence and transmission dynamics in β-thalassemics in India was determined
- Patient age and gender were key determinants for achieving spontaneous clearance
- HCV strains 3a and 1b were most prevalent in β-thalassemics in this region
- Uncommon HCV strains such as 3i and 3g were also isolated in this region
- Phylogeographic analysis mapped their transmission route via neighboring countries

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract:

Post transfusion acquired HCV infection is common in high risk group individuals such as β-thalassemia patients who depend on regular blood transfusions. This study was conducted to determine the prevalence, genotype distribution, transmission dynamics and correlation of patient's age and gender with spontaneous clearance of HCV in multi-transfused β-thalassemia patients, in the state of West Bengal, India. HCV RNA was first detected, using RT-PCR, from serum samples collected from HCV seropositive β-thalassemic individuals over five years. Viral RNA was further sequenced, and genotyped, using NCBI genotyping tool, for phylogenetic and phylogeographic studies using MEGA-X and BEAST package 1.10.0, respectively. Out of 777 HCV seropositive β-thalassemic patients, 503 (64.73%) were positive for HCV RNA, while 274 (35.26%) had spontaneously cleared the virus. Female thalassemic patients and patients belonging to ages 10-14 years, had a higher chance of spontaneously clearing the virus. The most prevalent HCV genotype in this study population was 3a (77.61%) followed by 1b (11.94%). Phylogeographic analyses revealed that the 3a strains migrated from countries like Pakistan, China Myanmar and Thailand whereas the 1b strains transmitted from Bangladesh and Myanmar. The prevalence of HCV infection is very high among Indian β-thalassemic patients, which necessitates a critical look into the prevailing transfusion practices and requires implementation of more rigid donor screening criteria to decrease the rate of transfusion transmitted HCV infection, especially in the highly susceptible thalassemic patients. The use of more sensitive NAT based assays for HCV detection in donor blood is a compressing need of the hour.

Keywords:

Hepatitis C Virus, Thalassemia, Prevalence, Genotype Distribution, Phylogenetic Study, Phylogeographic Analyses

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Compliance with Ethical Standards

Disclosure of potential conflicts of interest*:*

The authors declare they have no conflict of interest.

Funding:

Funding was provided by the Indian Council of Medical Research (ICMR), India, and Department of Science and Technology (DST), Government of West Bengal, India.

Acknowledgements:

We would like to acknowledge all the past and present members of the Sadhukhan lab: Kallol Saha, Rushna Firdaus, Upasana Baskey, Priya K. Verma, Sagnik Bakshi, Raina Das, Ronita De, Sudipta K. Biswas and Maya Halder for all their guidance and assistance with this project. We would like to thank the patients and their families for their kind cooperation and participation to perform this study. We would also like to thank our collaborative institutes for their help and support.

Research involving Human Participants and/or Animals:

Compliance with ethical standards: The study was designed and conducted as per ethical guideline of Helsinki,1964 (and its later amendments)

Ethical approval: All procedures performed in studies involving human participants were in accordance with the 1964 Helsinki declaration (and its later amendments) and was approved by the Institutional Ethical Committee, Indian Council of Medical Research-National Institute of Cholera and Enteric Diseases (ICMR-NICED), Kolkata.

Informed consent:

Informed consent was obtained from all individual participants included in the study.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

1. INTRODUCTION

HCV is an enveloped, positive sense, ssRNA virus of Hepacivirus genus and *Flaviviridae* family, that can cause liver cirrhosis and hepatocellular carcinoma (HCC). HCV mainly transmits through blood, blood products and body fluids infecting an estimated 200 million people, globally [1]. After leaving the bloodstream, HCV enters the liver and replicates in hepatocytes, resulting in acute or chronic liver infection. Based on viral genetic diversity, HCV has been classified into seven genotypes (1-7) and over 67 subtypes [2]. Hence, in this era of Direct Acting Antivirals (DAAs), it is extremely important to know the genotype of an infected patient to predict genotype specific DAA treatment [3]**.** HCV genotypes differ from each other not only in nucleotide sequences but also in geographical distribution and mode of transmission [4]**.**

β-thalassemia is an inherited autosomal recessive disorder that arises from reduced or lack of synthesis of the beta globin chain of hemoglobin [5]**.** β-thalassemia is prevalent all over the world; however, the frequency is higher in the Mediterranean region, the middle east, the Indian subcontinent, and Southeast Asian countries [6]. Approximately one-tenth of the world's thalassemic patients are born in India every year, of which approximately 1-3% in southern India and 3-15% in northern India are carriers [7]**.** Since regular blood transfusion is the major treatment option for multi-transfused β-thalassemia patients, these patients are at a higher risk of acquiring transfusion transmitted infections (TTI) by HIV, HBV and/or HCV, post transfusion, of which HCV infection is the major TTI in these patients in India (Umer and Iqbal, 2016). Although, the risk of acquiring TTIs has significantly decreased since the implementation of preventive screening measures of donor's blood in 2002, thalassemic patients are still at a higher risk of acquiring post transfusion HCV , if the transfused blood was collected during the donor's seronegative "window" period [8].

Genomic diversity of HCV in thalassemia patients depends on geographical localization. The most prevalent HCV genotype in thalassemic patients in Pakistan [9] and Iran [10] is genotype 3, genotype 4 predominates in Jordan [11], genotype 1 in Italy [12] and Greece [13]. In India, genotype 3 is prevalent in the north, east and west and genotype 1in the south [14]. Although, a few studies on the prevalence and genotypic distribution of HCV in thalassemia patients, were conducted by us in and around the capital city of West Bengal, Kolkata, (Biswas. et al., 2016), there was still a lack

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

of a more comprehensive study on the seroprevalence and genomic diversity of HCV in β-thalassemic individuals from India. This study was undertaken to highlight the seroprevalence, genomic diversity, migration and evolution dynamics of HCV, within the β-thalassemia patient population, in the Indian state of West Bengal from 2014-2018.

2. MATERIALS AND METHODS

2.1. Ethical Statement: Written informed consent was obtained from the patients before including them for this study. This study protocol complied with the Helsinki Declaration of 1975, amended in 2013, and was approved by the Institutional Ethical Committee, Indian Council of Medical Research-National Institute of Cholera and Enteric Diseases (ICMR-NICED), Kolkata, India.

2.2. Study Design: A total of 777 β-thalassemic patients seropositive for anti-HCV Ab were included in this study. HCV seropositivity was preliminarily measured from patient sera using a Tri-Dot assay and confirmed by ELISA (BioRad). These thalassemic patients received at least 25 units of blood transfusion, from January 2014 to December 2018, did not have history of any other viral infection or medical conditions, and were referred from 10 collaborating transfusion centers (TCs). These 10 TCs were named as TC 1-10 and are located around eight southern districts namely, Bankura, East Burdwan, West Burdwan, East Midnapur, West Midnapur, Howrah, Hooghly, and Kolkata, in the east Indian state, West Bengal. All laboratory work were done at ICMR-NICED, Kolkata. Blood samples were collected by venipuncture in clot vials, followed by collection of demographic data and complete patient history. HCV RNA positive patients were classified into seven age groups (AG), 1 to 7, depending on their age namely, AG-1 (2-5 years), AG-2 (6-9 years), AG-3 (10-14 years), AG-4 (15-19 years), AG-5 (20-24 years), AG-6 (25-29 years), and AG-7 (30 years and above). Patient age ranged from 2 to 35 years with a mean (±SD) of 11.5 ± 7.5 years. Spontaneous clearance was defined as presence of no or very low HCV RNA, (viral load >50 IU/ml) and was detected by RT-PCR. Patients who achieved spontaneous clearance in this study, were followed up for 6-12 months, after initial detection of infection and viral RNA clearance was confirmed by RT-PCR. Thalassemic patients receiving less than 25 units of blood or any kind of HCV treatment (interferons or antivirals) were not included in this study.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2.3. *Calculations*: Means and standard deviations were calculated using Excel. Statistical analyses for comparison between two groups was performed using a 'N-1' Chi-squared test. p value of <0.05 was considered as significant and was represented by "*" wherever indicated.

2.4. Detection of HCV RNA: Viral RNA was extracted from all seroreactive individuals using QIAamp viral RNA mini kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. Briefly, viral RNA was extracted from 140µl serum, eluted in 50µl elution buffer and stored at -80ºC for further use. Detection of HCV viral RNA was done by nested RT-PCR based on 5´UTR of HCV genome as described previously (Saha et al., 2014) [17]. Briefly, first round one tube RT-PCR was done in 20µl total reaction volume containing 2µl isolated RNA and second round nested PCR was performed in 25µl total volume using 2µl of 1st round RT-PCR product. A positive band at 256 bp in 1.5 % agarose gel stained with ethidium bromide was observed using a gel documentation system (BioRad, USA). Quantitative HCV RNA was estimated using a Qiagen quantitative real time RT-PCR (RT-qPCR) kit (QuantiFast Pathogen RT-PCR plus IC Kit). The HCV primer and probe sequences were directed against the 5´ UTR of the HCV genome [18] using the 4th WHO International Standard for HCV (NIBSC code 06/102) as standard.

Nested RT-PCR amplified amplicons of partial HCV core gene (405 bp) [16] were gel purified and directly used for DNA sequencing analyses in an automated DNA sequencer, model 3130XL (ABI, USA) using Big Dye terminator 3.1 kit (Applied Biosystems, USA). Informed consents for sample sequencing could be obtained from 335 (66.6%) of the 503 HCV RNA positive patients. The sequences were aligned and edited using the Benchling software (www.benchling.com), NCBI genotyping tool (www.ncbi.nlm.nih.gov/projects/genotyping/) and Clustal omega program (www.ebi.ac.uk/Tools/msa/clustalo).

2.5. Phylogenetic analysis: Phylogenetic analysis was performed using 80 representative partial core sequences from our HCV RNA positive samples. To investigate the evolutionary linkage among lab strains and reference strains, partial core sequences of 57 lab isolated 3a strains (GenBank Accession Numbers, MN635186-MN635242), 18, 1b strains (GenBank Accession Numbers, MN617343-MN617360) and 5, 3b strains (GenBank Accession Numbers, MN605946-MN605950), were aligned with HCV reference strains using Molecular Evolutionary Genetics Analysis

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

tool (MEGA-X) [19]. The evolutionary history was inferred using the Neighbor-Joining (NJ) method [20]**.** The optimal tree with the sum of branch length = 3.20301776 is shown in Figure 3b. The tree was drawn to scale, with lengths of branches in the same units as those of the evolutionary distances used to understand the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method [21] and are in the units of the number of base substitutions per site. This analysis involved 104 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 919 positions in the final dataset.

2.6. Phylogeographic analysis: To understand the transmission dynamics and HCV genotype distribution with evolutionary linkages in β-thalassemia population, a phylogeographic tree was constructed using Bayesian Evolutionary Analysis Sampling Trees (BEAST), package 1.10.4 (http://beast.community/programs). HCV core sequences from the following countries, India, China, Pakistan, Bangladesh, Myanmar, Thailand, Japan, Vietnam, Iran, Australia were downloaded from ViPR (Virus Pathogen Resource) database (www.viprbrc.org/brc/home.spg?decorator=vipr), HCV sequence database (https://hcv.lanl.gov/content/index) and NCBI. 352 bp (371 nt to 723 nt of the H77-accession number NC-004102) of the sequenced core region was used for analysis. The General Time Reversible (GTR) model of nucleotide substitution was used with gamma + invariant sites distribution. For the clocks setting, uncorrelated relaxed clock type was used [22] along with the Bayesian Skyline Coalescent model [23]. Markov Chain Monte Carlo (MCMC) analyses was performed with Path Sampling (PS)/stepping-stone (SS) analysis [24]. To ensure achievement of Effective Sampling Sizes (ESS) in TRACER (http://beast.community/tracer) [25], MCMC was run for 200 million states and sampled every 10000 states. The result of BEAST analyses was evaluated in TRACER. Tree Annotator (https://beast.community/treeannotator) and Fig-Tree (https://beast.community/figtree) software were used to construct the final tree.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

3. RESULTS

3.1. Distribution of HCV viremia in the ten transfusion centers:

The 777 HCV seropositive serum samples were collected from multi-transfused β-thalassemia patients at either of the 10 TCs located around eight southern districts of West Bengal, India as indicated in **(Supplemental Figure 1)**. HCV RNA was detected in a total of 64.73% (n=503) samples by qualitative nested RT-PCR; results were confirmed and viral load was measured by quantitative Real Time PCR. Prevalence of HCV RNA as observed in the 10 TCs was 81.25% (TC-1, n=26), 70.27% (TC-2, n=26), 70.58% (TC-3, n=24), 64.20% (TC-4, n=174), 58.51% (TC-5, n=55), 56.75% (TC-6, n=21), 82.27% (TC-7, n=65), 60.52% (TC-8, n=23), 52.56% (TC-9, n=41) and 62.33% (TC-10, n=48). The distribution of HCV viremia among the different TCs is depicted in **(Figure 1)**.

3.2. Co-relation of patient age and gender with spontaneous clearance:

Gender based studies revealed that out of the 503 HCV RNA positive samples, 62.82% (n=316) were males and 37.17% (n=187) were females **(Table 1)**. The highest percentage of patients, both males (34.49%, n=109) and females (29.94%, n=56) belonged to AG-3 (10-14 years) **(Figure 2a)**. In majority of cases the highest viral load (>100,000 IU/ml) was found in individuals in AG-5 and AG-6 (20-29 years) whereas AG-2 and AG-3 (6-14 years) had lower viral loads (50 IU/ml to 10,000 IU/ml) (data not shown). Our qRT-PCR data revealed that 35.26% (n=274) of our study population had spontaneously cleared the virus. Out of these 274 individuals, 51.82% (n=142) were males and 48.17% (n=132) were females **(Table 1)**. Interestingly, age groups 2 (Males: 27.46%, n=39; Females: 25.75%, n=34) and 3 (Males: 30.98%, n=44; Females: 32.57%, n=43), had significantly higher numbers of both male and female patients, who achieved spontaneous clearance, than individuals of the other age groups **(Figure 2b)** ($p<0.005$). We also found that HCV infected Indian β-thalassemic female patients had a significantly higher probability of clearing the virus (41.37%) as compared to their male counterparts (31.00%) ($p<0.005$) **(Table 1)**.

3.3. HCV genotype distribution:

335 partial core gene sequence (405 bp) were amplified and sequenced for HCV genotyping and phylogenetic analyses. Sequence alignment with reference HCV strains revealed that 87.76% (n=294) patients of our study population were infected with HCV genotype 3 and 12.23% (n=41) patients were infected with HCV genotype 1. The circulating HCV subtypes that were detected in this region were: 1a, 1b, 3a, 3b, 3g, and 3i with subtype 3a being the

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

predominant subtype (77.61%, n = 260), followed by subtype 1b (11.94%, n=40) and 3b (8.65%, n=29), while subtypes 3g, 3i and 1a were found in only 3 (0.89%), 2 (0.59%) and 1 (0.29%) patient(s), respectively **(Figure 3a)**. Studies on the distribution of HCV genotypes in each TC individually revealed that HCV genotype 3a was detected in samples collected from all the 10 TCs (TC1-TC10) and was the most prevalent HCV strain among all these 10 centers. HCV strains 1b and 3b could be found in only seven (TC-1 to TC-7) and nine (TC-1 to TC-9) TCs respectively, whereas HCV strains 1a, 3g and 3i were detected from one (TC-6), two (TC-5, TC-9) and one (TC-7) TC(s) respectively **(Supplemental Figure 2)**. Population based DNA sequencing data from the partial core region of HCV genome showed no multiple HCV genotype infections in any of the HCV RNA positive β-thalassemic patients in our study population **(Figure 3b).** The sequences reported in this study were deposited in gene bank with accession numbers as mentioned in the materials and methods section.

3.4. Phylogeography:

To examine the specific pathway of transmission dynamics between India and other adjacent countries, a retrospective Bayesian dated phylogeographic analysis was performed using core gene sequences of HCV strains obtained from PubMed sources available online. The maximum clade credibility tree was constructed on the basis of partial core gene sequences for genotypes 3a and 1b. We found that the majority of HCV genotype 3a sequences isolated in our laboratory formed four distinct clusters resulting in the China clade, Myanmar clade, Thailand clade and India/Pakistan clade **(Figure 4a)** whereas our lab isolated strains of HCV genotype 1b were similar to isolates from Bangladesh and Myanmar **(Figure 4b)**.

4. DISCUSSION

HCV infection is one of the most common TTIs in multi-transfused β-thalassemic individuals causing serious health complications in this high-risk group population. The epidemiology of HCV infection in thalassemic individuals is not very well reported from India. Determination of HCV genotypes and subtypes is essential to know the source of HCV infection in a specified population and it is also required for the clinical management, therapeutic intervention and development of an effective HCV vaccine. This study was carried out over a period of five years (January, 2014-December, 2018) in 777 HCV seropositive thalassemic patient samples that were collected from a total of 10 transfusion centers, where patients from several districts of West Bengal such as Bankura, Birbhum,

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Burdwan, East Midnapur, West Midnapur, Hooghly, Howrah, Purulia, Murshidabad, Nadia, Kolkata, North 24 Parganas, and South 24 Parganas, received blood transfusion **(Supplemental Figure 1)**. The prevalence of active HCV infection (state of viremia) in some of these transfusion centers was as high as 81% **(Figure 1)**. These observations are very interesting because all the samples for this study were collected over a period of five years from different transfusion centers, that are located in different districts of West Bengal, yet there was a high prevalence of transfusion transmitted HCV infection at all 10 centers. This data suggests that TCs act as hotspots for transmitting HCV infection due to use of contaminated medical equipment. It can also be postulated that due to improper screening of donor's blood during the seronegative "window" period, serological tests such as Tri-Dot assays and ELISA, which are more commonly used for donor blood screening in the blood banks are not sufficient to detect the presence of active HCV infection in donor blood. Hence HCV infected blood is getting transfused to the thalassemic recipients. Thus, maintenance of strict transfusion protocols using properly sterilized or disposable medical equipment and more sensitive NAT based RT-PCR screening methods are essential to check donor's blood for probable infections before it is used for transfusion to prevent the transmission of HCV in high-risk group population.

Our data revealed that β-thalassemic patients within the age group of 10-14 years had a higher prevalence of HCV viremia **(Figure 2a)**. Interestingly, this particular age group had the highest number of spontaneous clearance cases suggesting that patients whose age is between 10-14 years had a higher chance of clearing the infection spontaneously as compared to patients of other ages **(Figure 2b)**. We also found that there was a correlation between age and viral titer as patients with lower age (6-14 years) had low HCV viral titer as compared to patients with higher age (20-29 years) suggesting that low age thalassemic patients have a higher probability of clearing the infection, than older patients. This could be because of the increase in serum ferritin levels in older thalassemic patients due to repeated blood transfusions, that facilitates HCV replication [26] and leads to faster progression to liver cirrhosis and HCC [27].

Guy and Peters have previously reported that women can clear acute HCV infection at a higher rate than men [28]**.** Our data provides evidence for the first time that in HCV infected multi-transfused β-thalassemic Indian patients, thalassemic females were more likely to spontaneously clear the virus than thalassemic males **(Table 1)**. This difference in HCV clearance rate based on gender is still an unresolved issue, but there are a few plausible explanations. Since, women have a slower rate of liver disease progression including liver cirrhosis and HCC [29], it is not surprising that the burden of HCV disease complications and the rate of infection clearance is lower in thalassemic

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

males as compared to thalassemic females**.** Secondly, in most viral infections, basal immune responses are higher in females than in males and low doses of estrogen have been suggested to modulate antiviral innate and adaptive immune responses [30]. However, in addition to these factors, several other yet to be determined genetic and immunologic factors can play a key role in favoring HCV clearance in the thalassemic females of our study population.

HCV genotypes that we report in this study are 1a, 1b, 3a, 3b, 3g and 3i **(Figure 3a, 3b)** of which genotypes 3g and 3i have not been previously reported from this region and this is the first time that these two HCV genotypes were found circulating in β-thalassemic patients of West Bengal. Another interesting finding is the higher predominance of HCV genotype 3 (subtype a) in the β-thalassemic patients, irrespective of the fact that the transfusion centers from where these patients received blood transfusions were all located in different districts. Although the predominance of HCV genotype 3 (63.85%) in the Indian population have been reported earlier [31] but our study reports that the prevalence of HCV genotype 3 among the Indian β-thalassemic patients is even higher (87.76%). Several factors are known to influence HCV replication such as host immunity, age, hepatocellular damage, whether the infection is acute or chronic and these factors might provide a replicative advantage to one HCV genotype in a specific patient population than other genotypes [32]. Although a more detailed analyses is required to prove this but from our results it can be preliminarily speculated that HCV genotype 3a has a replication priority over other HCV genotypes during the natural course of infection in the β-thalassemic Indian population.

Phylogeographic analysis of HCV strains assists in monitoring the virus for molecular tracing, ancestral studies, genetic assessments, clinical intervention and for treatment decision. Hence, we performed phylogeographic analyses of the two most commonly found HCV genotypes among β-thalassemia patients in West Bengal, 3a and 1b, to obtain information about the hierarchal relationships and genetic evolution. However, we could not perform phylogeographic analyses of HCV strains 1a, 3i and 3g as we did not isolate sufficient numbers of these strains from our study population to undertake any such analytical study. We found that the majority of the HCV 3a isolates from West Bengal have sequence similarities with HCV strains from neighboring countries like Pakistan, China, Myanmar and Thailand **(Figure 4a)** whereas majority of the HCV 1b isolates of this region had sequence similarities to HCV strains of Bangladesh and Myanmar **(Figure 4b)**, suggesting that the point of origin of the circulating HCV strains 3a and 1b can be traced back to the adjacent countries that share border with India.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

5. CONCLUSION

To summarize, this study provides the first multi-district, comprehensive and thorough analyses of HCV prevalence, viremia and genomic diversity in β-thalassemia patients in West Bengal, India, reporting that the two most prevalent HCV strains circulating in β-thalassemia patients in this part of the country are 3a and 1b. We provide evidence supporting the presence of additional HCV strains 3i and 3g, which are not commonly reported from this region. Our study also indicated that young adults between the ages of 10-14 years and female β–thalassemic patients had a higher probability of spontaneously clearing the virus. Additionally, we report that our lab-isolated HCV strains 3a and 1b, migrated to West Bengal through transmission from neighboring countries, China, Pakistan, Myanmar and Bangladesh.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References:

1. Zeller H, Rabarijaona L, Rakoto-Andrianarivelo M, Boisier P. Prevalence of hepatitis C virus infection in the general population of Madagascar. *BullSocPatholExot*. 2017;90(1):3-5.

2. Smith DB, Bukh J, Kuiken C, et al. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource. *Hepatology*. 2014;59(1):318-327. doi:10.1002/hep.26744

3. Cavalcante LN, Lyra AC. Predictive factors associated with hepatitis C antiviral therapy response. *World J Hepatol*. 2015;7(12):1617-1631. doi:10.4254/wjh.v7.i12.1617

4. Pozzetto B, Memmi M, Garraud O, Roblin X, Berthelot P. Health care-associated hepatitis C virus infection. *World J Gastroenterol*. 2014;20(46):17265-17278. doi:10.3748/wjg.v20.i46.17265

5. Cao A, Galanello R. Beta-thalassemia. *Genet Med*. 2010;12(2):61-76. doi:10.1097/GIM.0b013e3181cd68ed

6. Vidja PJ, Vachhani JH, Sheikh SS, Santwani PM. Blood transfusion transmitted infections in multiple blood transfused patients of beta thalassaemia. *Indian J Hematol Blood Transfus*. 2011;27(2):65-69. doi:10.1007/s12288-011-0057-3

7. Joseph N, Pai S, Sengupta S, Bharadwaj S, Dhawan S, Khare K. A clinico-epidemiological study of thalassemia cases in India. *J Nat Sci Biol Med*. 2018;9(2):236. doi:10.4103/jnsbm.jnsbm_224_17

8. Biceroglu SU, Turhan A, Doskaya AD, et al. Probable hepatitis C virus transmission from a seronegative blood donor via cellular blood products. *Blood Transfus*. 2014;12(SUPPL.1):69-70. doi:10.2450/2013.0285-12

9. Umer M, Iqbal M. Hepatitis C virus prevalence and genotype distribution in Pakistan: Comprehensive review of recent data. *World J Gastroenterol*. 2016;22(4):1684-1700. doi:10.3748/wjg.v22.i4.1684

10. Ahmadi-Ghezeldasht S, Badiei Z, Sima HR, et al. Distribution of Hepatitis C Virus Genotypes in Patients with Major β-Thalassemia in Mashhad, Northeast Iran. *Middle East J Dig Dis*. 2018;10(1):35-39. doi:10.15171/mejdd.2017.88

11. Al-Sweedan SA, Jaradat S, Amer K, Hayajneh W, Haddad H. Seroprevalence and genotyping of hepatitis C virus in multiple transfused Jordanian patients with β-thalassemia major. *Turkish J Hematol*. 2011;28(1):47-51. doi:10.5152/tjh.2011.05

12. Di Marco V, Capra M, Angelucci E, et al. Management of chronic viral hepatitis in patients with thalassemia: Recommendations from an international panel. *Blood*. 2010;116(16):2875-2883. doi:10.1182/blood-2009-11-248724

13. Triantos C, Konstantakis C, Tselekouni P, Kalafateli M, Aggeletopoulou I, Manolakopoulos S. Epidemiology of hepatitis C in Greece. *World J Gastroenterol*. 2016;22(36):8094-8102. doi:10.3748/wjg.v22.i36.8094

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

14. Satsangi S, Chawla YK. Viral hepatitis: Indian scenario. *Med J Armed Forces India*. 2016;72(3):204-210. doi:10.1016/j.mjafi.2016.06.011

15. A. B, K. S, R. F, et al. Prevalence of ANTI-HCV, HBSAG, HIV among multi-transfused thalassemic individuals and their socio-economic background in Eastern India. *Asian J Pharm Clin Res*. 2016;9(1):290-294. http://innovareacademics.in/journals/index.php/ajpcr/article/download/9134/3622%0Ahttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed17&NEWS=N&AN=607582873.

16. Saha K, Biswas A, Firdaus R, Sadhukhan PC. A novel nested RT-PCR method for Rapid HCV Detection and Genotyping. 2014;(APRIL):1-12.

17. Biswas A, Firdaus R, Gupta D, et al. Interferon λ3 gene (IL28B) is associated with spontaneous or treatment-induced viral clearance in hepatitis C virus–infected multitransfused patients with thalassemia. *Transfusion*. 2017;57(6):1376-1384. doi:10.1111/trf.14070

18. Firdaus R, Biswas A, Saha K, et al. Impact of host IL28B rs12979860, rs8099917 in interferon responsiveness and advanced liver disease in chronic genotype 3 hepatitis C patients. *PLoS One*. 2014;9(6):1-8. doi:10.1371/journal.pone.0099126

19. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35(6):1547-1549. doi:10.1093/molbev/msy096

20. Saitou N, Nei M, Kumar S, et al. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35(4):1547-1549. doi:10.1093/molbev/msy096

21. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci*. 2004;101(30):11030-11035. doi:10.1073/pnas.0404206101

22. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006;4(5):699-710. doi:10.1371/journal.pbio.0040088

23. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol*. 2005;22(5):1185-1192. doi:10.1093/molbev/msi103

24. Saha K, Firdaus R, Biswas A, et al. Transmission dynamics of hepatitis C virus among intra venous drug users in the border state of Manipur, India. *Infect Genet Evol*. 2014;24:57-67. doi:10.1016/j.meegid.2014.03.008

25. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol*. 2018;67(5):901-904. doi:10.1093/sysbio/syy032

26. Zou DM, Sun WL. Relationship between hepatitis C virus infection and iron overload. *Chin Med J (Engl)*. 2017;130(7):866-871. doi:10.4103/0366-6999.202737

27. Marcus E-L, Tur-Kaspa R. Chronic Hepatatis C Virus Infection in Older Adults. *Clin Infect Dis*. 2013;41(11):1606-1612.

28. Guy J, Peters MG. Liver disease in women: The influence of gender on epidemiology, natural history, and patient outcomes. *Gastroenterol Hepatol*. 2013;9(10):633-639.

29. Baden R, Rockstroh JK, Buti M. Natural history and management of hepatitis C: Does sex play a role? *J Infect Dis*. 2014;209(SUPPL. 3):81-85. doi:10.1093/infdis/jiu057

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

30. Ruggieri A, Gagliardi MC, Anticoli S. Sex-dependent outcome of hepatitis B and C Viruses infections: Synergy of sex hormones and immune responses? *Front Immunol*. 2018;9(OCT):1-7. doi:10.3389/fimmu.2018.02302

31. Christdas J, Sivakumar J, David J, Daniel HDJ, Raghuraman S, Abraham P. Genotypes of hepatitis C virus in the Indian sub-continent: A decade-long experience from a tertiary care hospital in South India. *Indian J Med Microbiol*. 2013;31(4):349-353. doi:10.4103/0255-0857.118875

32. Yang Z, Yu Y, Zhang H, et al. Replication priority of hepatitis C virus genotype 2a in a Chinese cohort. *Acta Pharm Sin B*. 2014;4(4):266-269. doi:10.1016/j.apsb.2014.06.011

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

TABLES

Table 1: Gender based viral RNA clearance in multi-transfused β-thalassemic patients.

	RNA+	**RNA-**	**Total**	**RNA Clearance (%)**
Males	316	142	458	31.00
Females	187	132	319	41.37
Total	503	274	777	
***p* value**				<0.005

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure Legends:

Fig. 1: HCV RNA prevalence in multi-transfused thalassemic individuals

Samples were collected from 10 transfusion centers (TC-1 to TC-10). Percentage of HCV RNA positive samples from each of the 10 TCs along with the number of total HCV RNA positive samples collected from each center. TC: Transfusion Centre.

Fig 2a: Age and gender-based prevalence of active HCV infection among β-thalassemic patients

Percentage of HCV RNA positive samples along with the total number of HCV RNA positive samples in each age group (AG) in the male and female population.

Fig. 2b: Age and gender-based spontaneous clearance of HCV infection among β-thalassemic patients

Percentage of HCV RNA negative samples along with the total number of HCV RNA negative samples in each age group in the male and female populations. AG: Age group. ∗ denotes $p<0.005$.

Fig. 3a: Genetic distribution of different HCV genotypes and subtypes

The genetic distribution of the circulating HCV strains along with their respective percentages as identified in β-thalassemic individuals.

Fig. 3b: Phylogenetic tree was constructed from partial core gene sequences isolated from HCV RNA positive thalassemia patients

The reference sequences used are Genotype-1a (AF511950, M62321); Genotype-1b (D11355, AY587016, EF032892, EU782828, D90208); Genotype-1c (D14853, AY051292); Genotype-2a (D00944, AB047639); Genotype-2b (AF238486, D10988); Genotype-3a (D17763, JN714194, X76918, D28917, AF046866); Genotype-3b (JQ065709, D49374); Genotype-4a (DQ418789); Genotype-5a (Y13184); Genotype-6a (Y12083); Genotype-7a (EF108036). The geometric symbols provided next to the sequences denote their respective subtypes with respect to their reference sequences. The GenBank Accession numbers of the HCV lab isolates used for this image were: Genotype 3a (GenBank Accession Numbers, MN635186-MN635242), Genotype 1b (GenBank Accession Numbers, MN617343-MN617360) and Genotype 3b (GenBank Accession Numbers, MN605946-MN605950).

Fig. 4a: Phylogeographic analysis of HCV genotype 3a strain isolated from β-thalassemia patients

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The maximum clade credibility tree of the HCV genotype 3a was constructed based on the genomic region 371nt to 723nt of the H77 genome (NC-004102). The ancestral geographic states were reconstructed using Bayesian MCMC BEAST v1.10.4 package. The tree was constructed using reference sequences from China: KC844041, HQ639941, HQ639942, HQ912953; India: JN714194; Iran: KC285360, KC285362, KC285358; Pakistan: HQ108102, HQ108103, HQ108104, HQ108105; Japan: D17763, Bangladesh: JQ888224, JQ668320, JQ888231, Thailand: D14307, HQ229199, HM042020, HM042016; Myanmar: GU186941, AB290962, AB269327; Australia: AF046866. The local sequences were annotated with prefix Thal. The GenBank Accession numbers of the HCV 3a lab isolates used for this image were: Genotype 3a (GenBank Accession Numbers, MN35194, MN635196, MN635198, MN635201, MN635205, MN635206, MN635208, MN635211, MN635213, MN635220, MN635222, MN635225, MN635227, MN635228, MN635234, MN635235, MN635237, MN635239, MN635241, MN635242).

Fig. 4b: Phylogeographic analysis of HCV genotype 1b strain isolated from β-thalassemia patients

The maximum clade credibility tree of the HCV genotype 3a was constructed based on the genomic region 371nt to 723nt of the H77 genome (NC-004102). The ancestral geographic states were reconstructed using Bayesian MCMC BEAST v1.10.4 package. The tree was constructed using reference sequences from Japan: AB062255, AB077724, FJ607091, FJ607115, FJ607117; Vietnam: AB301749, DQ155495, DQ155498, FJ768880; Thailand: AF525890, EF543230, GQ913859, GQ913860, HM041986, HQ229114, HQ229043; Myanmar: AB269350, AB269352, GU186959, FJ607133; China: DQ777809, EU081379, EU081407, EU081423, EU081424; India: AY921165, KF181661, KF181668, KF181665, KF181670; Bangladesh: JQ668313, JQ668314, JQ888223. The local sequences were annotated with prefix Thal. The GenBank Accession numbers of the HCV 1b lab isolates used for this image were: Genotype 1b (GenBank Accession Numbers, MN617343-MN617357, MN17360).

Table 1: RNA clearance rate in multi-transfused β-thalassemic males vs females. *p* value was calculated by a 'N-1' Chi-squared test.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Supplementary Figure Legends:

Fig. 1: Map of West Bengal showing the 10 transfusion centers (TC)

Illustrative map of the Indian state of West Bengal showing the ten TCs as red dots located around eight districts, namely, Bankura, East Burdwan, West Burdwan, East Midnapur, West Midnapur, Howrah, Hooghly, and Kolkata. Each TC was located in each district with the exception of Kolkata district which has three TCs.

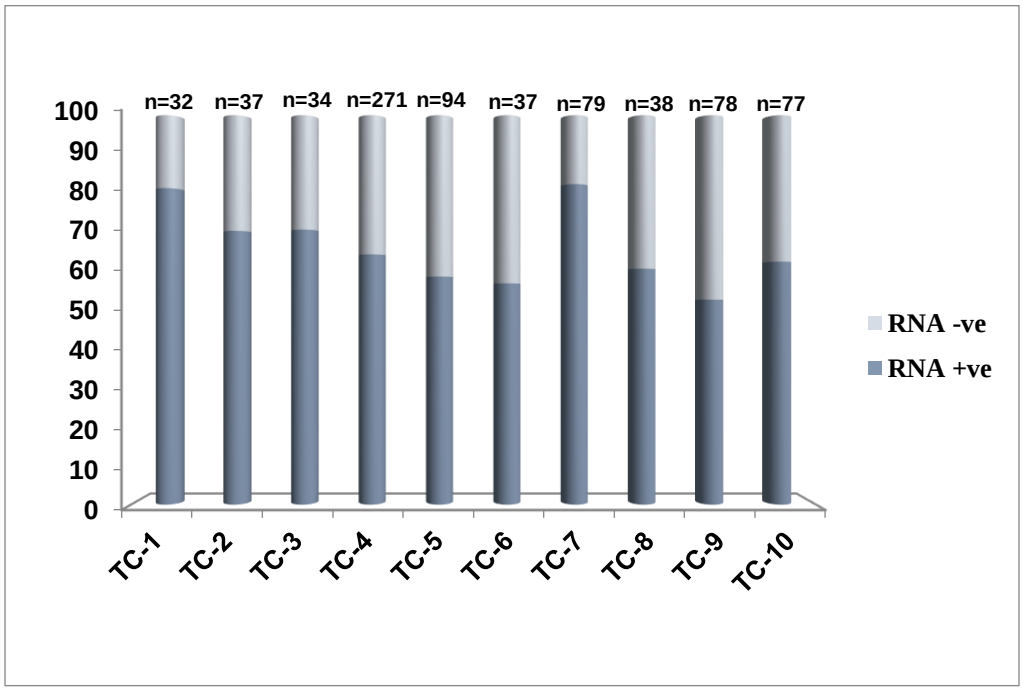
Fig. 2: Distribution of different HCV strains among β-thalassemic patients in the 10 Transfusion Centers

The genetic distribution of each of the 6 HCV strains 1a, 1b, 3a, 3b, 3g and 3i in the 10 TCs.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

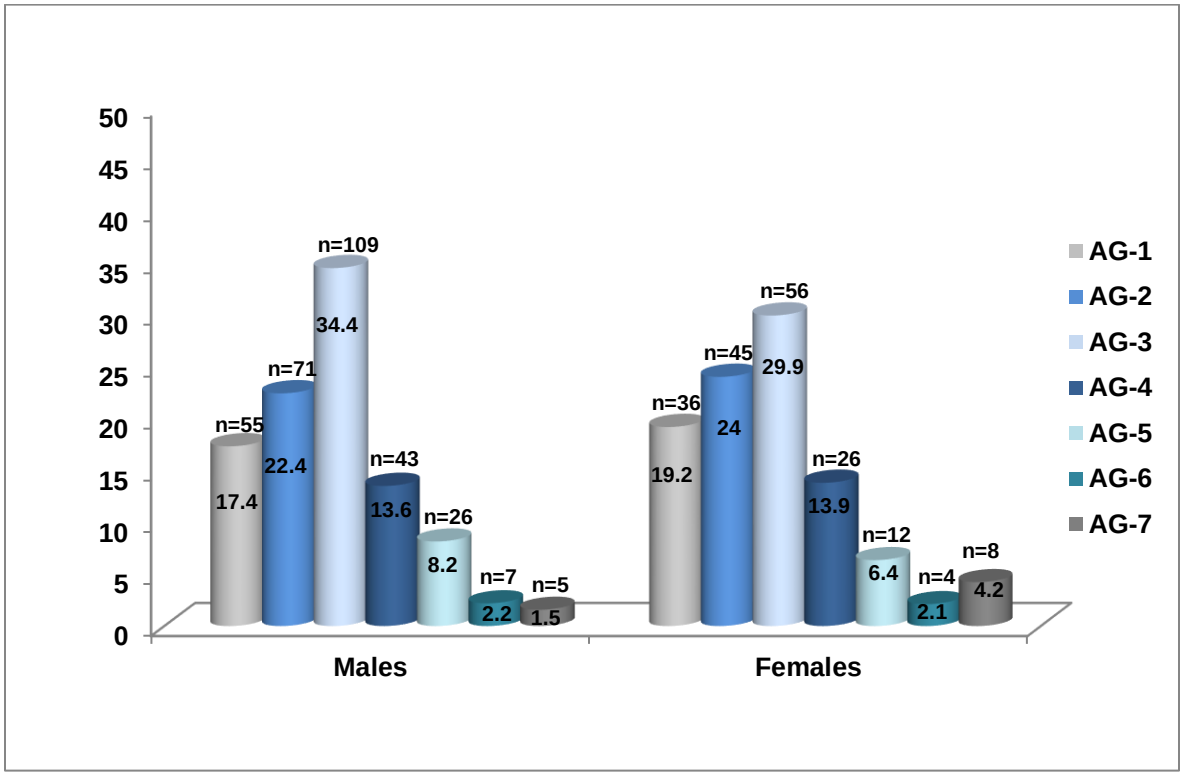
FIGURES

Figure 1: Prevalence of HCV RNA in HCV seropositive multi-transfused β-thalassemic individuals in the 10 Transfusion Centers



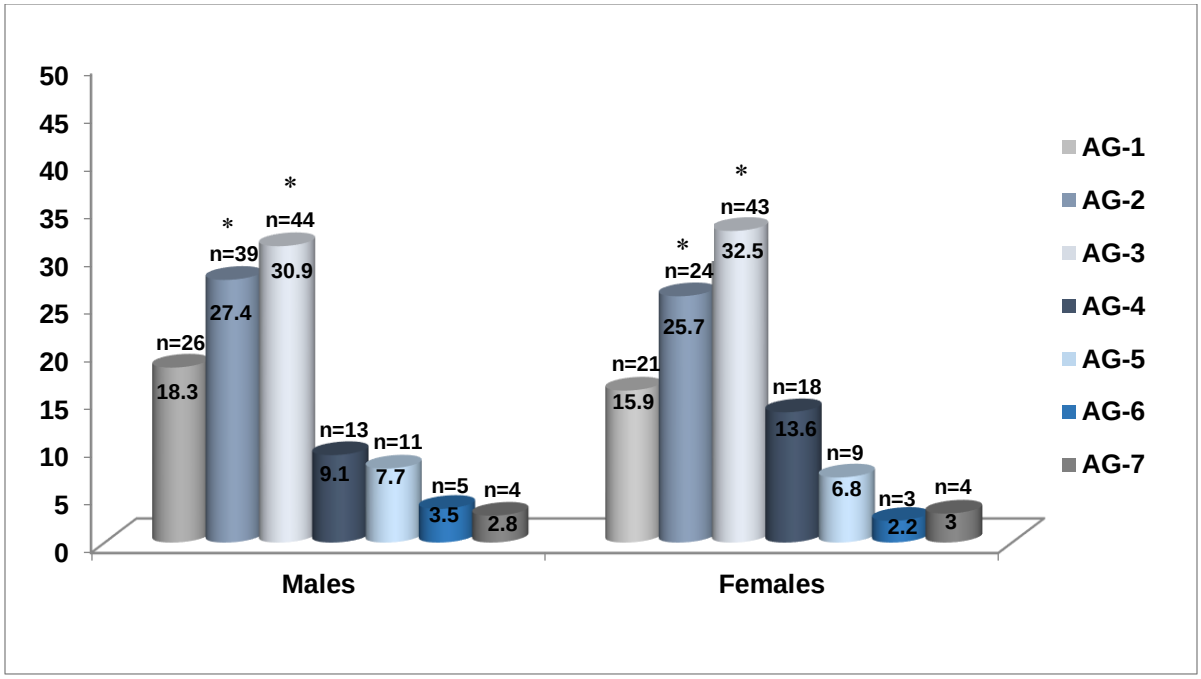
medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 2: (a): Gender and age-based prevalence of HCV RNA among β-thalassemic patients.



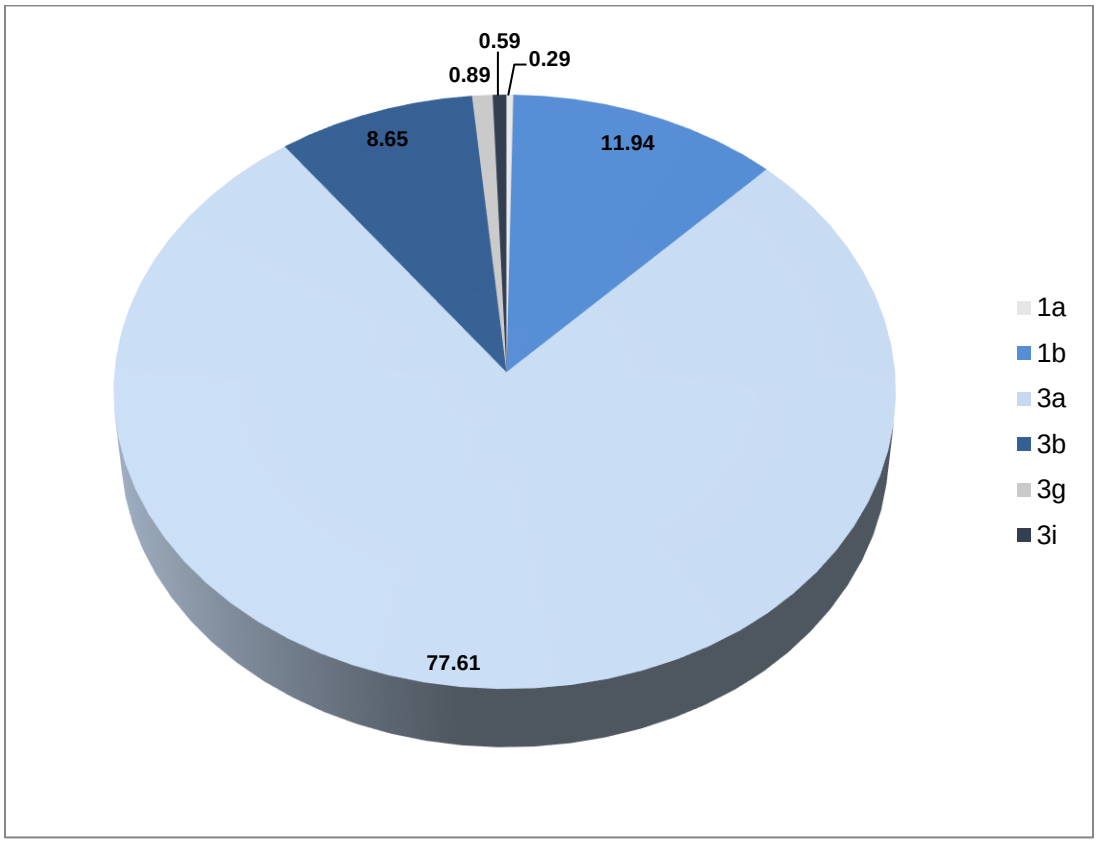
medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 2: (b): Gender and age-based spontaneous clearance of HCV RNA among β-thalassemic patients.



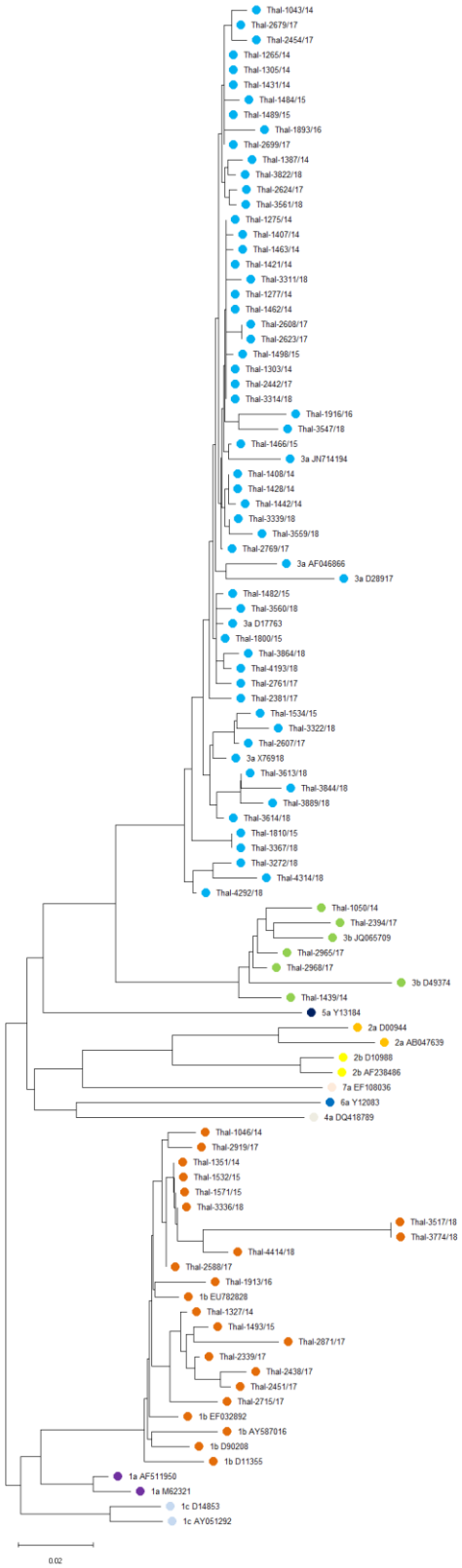
medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 3: (a): Distribution of different HCV genotypes and subtypes identified in our study population.



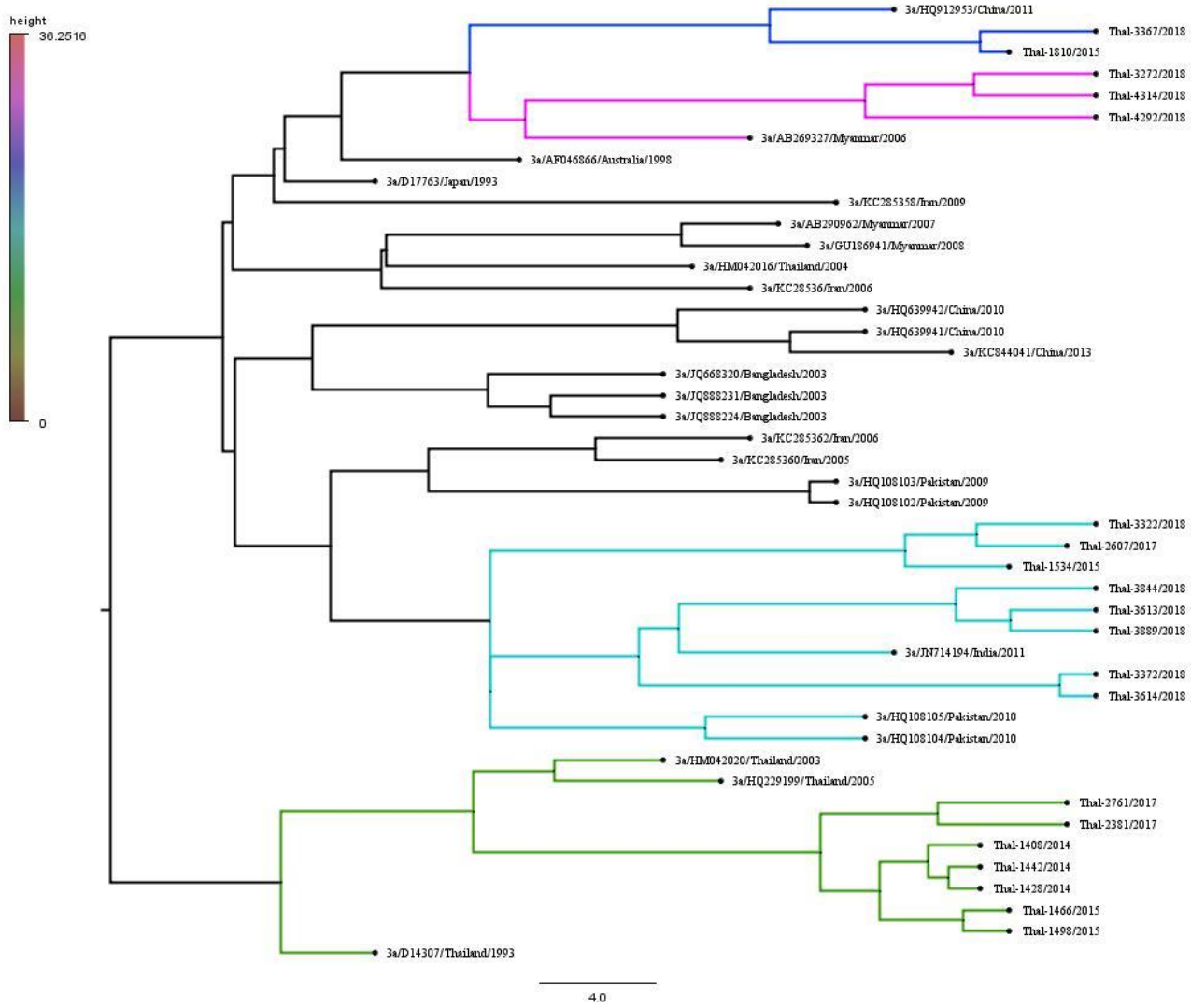
medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 3: (b): Phylogenetic tree obtained from 80 HCV partial core gene sequences.



medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 4: (a): Phylogeographic analysis of HCV genotype 3a strain isolated from β-thalassemia patients in this study.



medRxiv preprint doi: https://doi.org/10.1101/2020.05.06.20092049; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

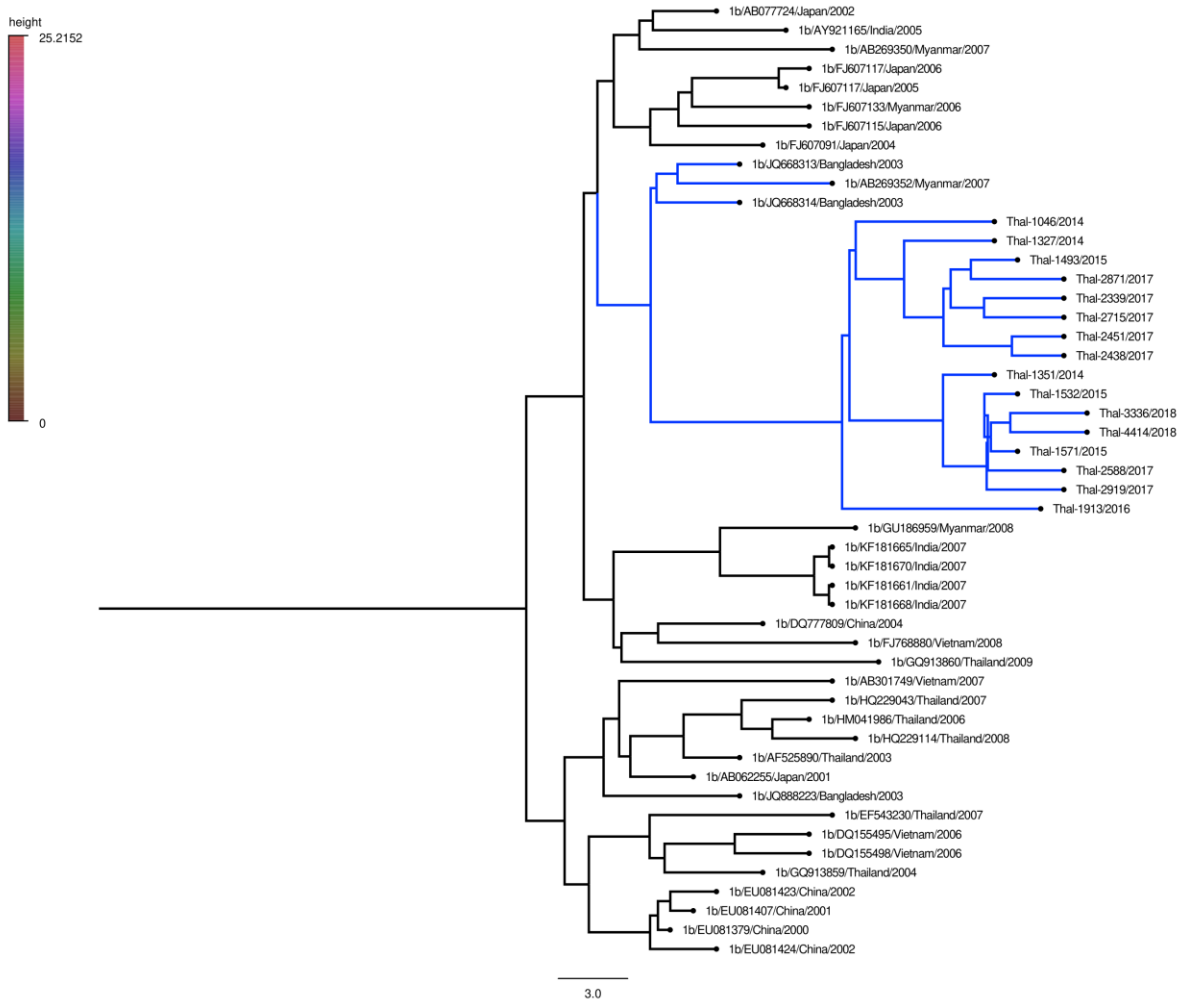



Figure 4: (b): Phylogeographic analysis of HCV genotype 1b strain isolated from β-thalassemia patients in this study.