

NGS evaluation of HCV strains from individuals co-infected with HIV

Zoya Dimitrova¹, Lilia Ganova-Raeva¹ and Ivailo Alexiev², Pavel Teoharov³,
Lili Punkova¹, Amanda Sue¹, Pavel Skums¹, and Yuri Khudyakov¹

¹ Centers for Disease Control and Prevention, MS-A33, Atlanta, GA 30329, USA
`izd7@cdc.gov`

² National Center of Infectious and Parasitic Diseases, National reference Laboratory
for HIV, Sofia, Bulgaria

³ National Center of Infectious and Parasitic Diseases, National Reference
Laboratory of Hepatitis, Sofia, Bulgaria

Abstract. The study was done with a cohort of 150 individuals identified in Bulgaria between 2010-2014 as HCV co-infected HIV carriers. We employed Next-generation sequencing (NGS) to investigate the diversity of the Hepatitis C strains circulating in among them. K-step network of all HCV haplotypes detected in these patients was created and the patients with samples from more than one time point were examined within the context of the network. The network revealed re-infection events, transmission links between cases and infections with multiple genotypes.

Keywords: Next-generation sequencing, viral quasi species

1 Introduction

Hepatitis C virus (HCV) infection is a major cause of liver disease in the world. It is estimated that 130 million people are infected with HCV globally [1]. Major risk factors for transmission of HCV are injection drug use (IDU) and male homosexual behavior (MSM). HIV-positive individuals are frequently co-infected with other pathogens. Owing to a common mode of transmission, HIV and HCV co-infections are frequent. The risk of infection with HCV is 4 to 10-fold higher than that of HIV [2]. Both viruses are genetically heterogeneous and exist as a large population of intra-host viral variants in each infected person. Assessing intra-patient viral genetic diversity is essential for understanding the evolutionary dynamics of viruses to assist in transmission control and treatment. NGS has emerged as powerful means for the acquisition of large amount of genetic data, but at the same time presents unique challenges in experimental design and computational data analysis [3].

The current study represents NGS data for a cohort of 150 individuals identified in Bulgaria between 2010-2014 as co-infected with HIV and HIV. Bulgaria has a population of 7 million and the cumulative number of HIV/AIDS cases reached 2077 in 2014 [4]. Although the the rate of the HCV infection among the general population in Bulgaria is relatively low (0.8-1.3%)[5], recent studies

[6] have found that the rate of HCV infections has increased among high risk behavior groups, including injection drug users (IDU), MSM (man that have sex with man), and incarcerated individuals.

2 Methods

NGS

Serum specimens were available from a total of 150 individuals from Bulgaria co-infected with HIV and HCV. A second follow up specimen was available from 14 of those individuals. The genetic target used in the study was Hyper Variable region 1 (HVR1) of the HCV envelope gene. All specimens were bar coded in a HVR1 nested reaction by using 10-mer barcodes. Specimens were pooled, and each pool was further bar coded with Adapters and 6-mer indexes (NEB Ultra DNA kit). The products were purified using Ampure XP (Agilent) and quantified on Tape Station Instrument (Agilent). The pools were normalized and combined to generate a 10pmol library for sequencing using the paired-end read protocol on Illumina MiSeq. All specimens were sequenced together in this fashion. The pools were de-multiplexed by indexes on the MiSeq instrument. Further de-multiplexing was done using CLC Genomics workbench 8.1 version software (CLC BIO, Qiagen). The raw-data files were trimmed for quality and length (≥ 260 nt), and pairs of reads were merged using the following parameters: Mismatch cost = 1, Minimum score = 10, Gap cost = 1, Maximum unaligned end mismatches = 2. Sequences were assigned to correct specimens from merged pairs by barcodes on both ends of the sequenced amplicon. The majority consensus was compared to the Sanger data from the corresponding patient. The NGS data were cleaned by the k-mer error correction (KEC) and empirical threshold algorithms using the parameters $k = 25$ and $i = 3$ [7].

One-step and k-step network construction

To study the dynamics of intra-host HCV evolution, we created a k-step network of intra-host variants sampled from all available NGS haplotypes. All unique haplotypes were aligned and the Hamming distance between each haplotype pair was calculated. Then we generated an one-step network where each unique haplotype was represented by a node and two nodes were connected by an edge if the distance between them was 1. The network consisted of several connected components. To join the components together, k-step network was constructed as follows: iteratively, for $k = 2, 3, \dots, k$, all pairs of haplotypes from different components with distance equal to k were found. They were linked by edges and the connected components were recalculated. These steps were repeated until a single connected component was formed. The resulting k-step network is equivalent to the union of all minimum spanning trees of the one-step network. The analysis and network visualization were performed using MATLAB R2015a (The MathWorks, Inc.) and GEPHI [8].

3 Results and Discussion

K-step network A k-step network was generated from the unique haplotypes (n=2018) obtained from all 123 cases that yielded NGS sequences. The network identified the following genotype and sub-type clusters: 1a, 1b, 2a, 3a and 4a.

Profiles of patients with follow up sample

The median age of the 14 patients was 25. Among the 14 cases, 12 had reported risk factor of IDU, three had been incarcerated, and one was a sex worker; 13 were females. The unique haplotypes (n=640) identified in the cases belonged to genotypes 1a, 1b and 3a. In all but one cases the Sanger consensus sequence was of same genotype as the most frequent haplotype. The second sample was taken between 1 - 18 months (6 months average) after the first one.

Quasi species evolution

Six of the patients were infected with a single HCV genotype: 1a (n=4, cases 582, 606, 1062, 1801) or 1b (n=2, cases 898, 1342). Figure 1 illustrates the k-step network for case 898 presented in the context of all detected haplotypes; blue nodes represent the haplotypes detected in the first time point and red nodes depict haplotypes sampled in the second time point. Several cases maintained a mixed infection with two genotypes, illustrated in Figure 2 for case 252. Three cases (1156, 1323, 1514) acquired super-infection with a different HCV strain, resulting in change of the predominant genotype. Figure 3 illustrates this with for case 1514, who acquired super infection with new genotype 3a.

Transmission links

Using the earlier developed algorithm [9], we detected 2 clusters of transmission; namely, one cluster of 1291, 1111 and 1313 and the other of 1514, 1156, 259, 596 and 582. The links between 1514 and 1156 were found for both genotypes 1a and 1b.

In most cases (582, 898, 1062, 1342, 1801), where a mono infection was maintained, a significant level of divergence could be observed. The same was true in some of the mixed infection cases (e.g. 252, 1111).

Conclusion

NGS provides valuable in-depth information about the complexity of an HCV infection. NGS is indispensable tool for transmission tracking and particularly in cases of IDU, where there are multiple exposures and opportunities for reinfection and super infection with a new strain. The information gained by NGS is not accessible by Sanger consensus sequencing alone or by serology testing. This is the first report describing HIV carriers co infected with HCV in this population.

References

1. Hepatitis C fact sheet, WHO, Sept 2015; <http://www.who.int/mediacentre/factsheets/fs164/en/>
2. Basnayake SK, Easterbrook PJ.,2016, Wide variation in estimates of global prevalence and burden and death estimates from chronic HBV, HCV and coinfection with HIV in literature, J Viral Hepat. 2016 Mar 30. Review.

3. Beerenwinkel, N., et al., 2012, Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. *Front Microbiol* 3: 329.
4. Ciccozzi M, Zehender G, Cento V, Lo Presti A, Teoharov P, Pavlov I, et al. Molecular analysis of hepatitis C virus infection in Bulgarian injecting drug users. *J Med Virol.* 2011;83(9):1565-0.
5. 2014, ECDC/WHO, HIV/AIDS Surveillance in Europe. Available from: http://ecdc.europa.eu/en/publications/surveillance_reports/HIV_STI_and_blood_borne_viruses/Pages/hiv_aids_surveillance_in_Europe.aspx
6. European Centre for Disease Prevention and Control. Hepatitis C surveillance in Europe 2013. Stockholm: ECDC; 2015. Available from: <http://ecdc.europa.eu/en/publications/Publications/epatitis-c-surveillance-in-europe-2013.pdf>
7. Skums P, Dimitrova Z, Campo DS, Vaughan G, Rossi L, Forbi JC, Yokosawa J, Zelikovsky A, Khudyakov Y. 2012. Efficient error correction for next-generation sequencing of viral amplicons. *BMC Bioinformatics* 13 Suppl 10:S6.
8. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: *International AAAI Conference on Weblogs and Social Media.* (San Jose, CA, USA.)
9. Campo DS, Dimitrova Z, Yamasaki L, Skums P, Lau DT, Vaughan G, Forbi JC, Teo CG, Khudyakov Y. 2014. Next-generation sequencing reveals large connected networks of intra-host HCV variants. *BMC Genomics* 15 Suppl 5:S4.

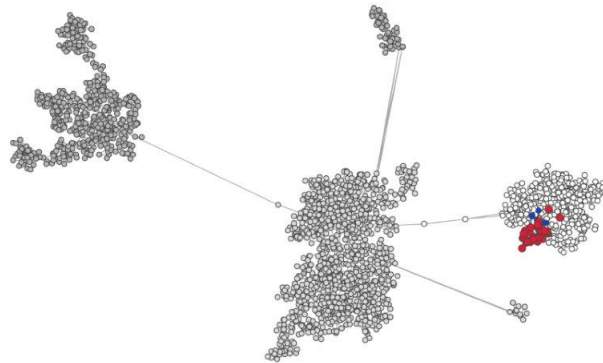


Fig. 1. K-step network of all identified haplotypes. Case 898 is highlighted as an example of mono infection; blue nodes - first time point; red nodes - second time point

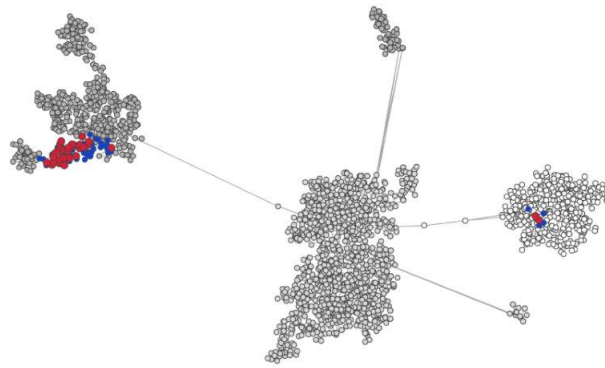


Fig. 2. Case 252 as example of an infection with two different genotypes, 3a and 1b, in the context of the K-step network; blue nodes - first time point; red nodes - second time point

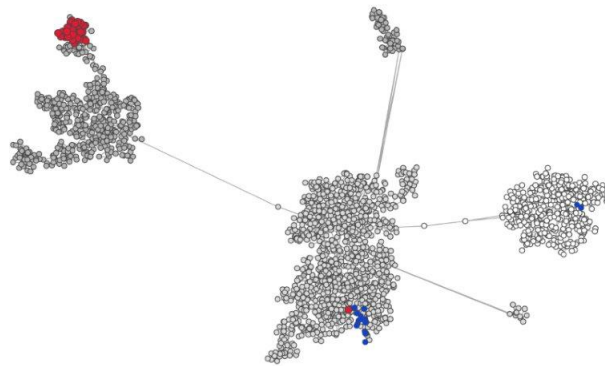


Fig. 3. Case 1514 in the context of the K-step network. Infection with two different genotypes, 1a and 1b, in the first time point and a super infection with third genotype-3a in the second time point ; blue nodes - first time point; red nodes - second time point.