Genetic diversity, haplotype analysis, and risk factor assessment of hepatitis a virus (HAV) isolates from the West Bank, Palestine during the period between 2014 and 2016

Kamal Dumaidi[1*], Hayah Qaraqe [4], Amer Al-Jawabreh[1,2], Rasmi Abu-Helu[3], Fekri Samarah [1], Hanan Al-Jawabreh[2]

[1]Department of Medical Laboratory sciences, Faculty of Allied Health Sciences, Arab American University, Palestine

[2] Leishmaniases Research Unit, Jericho, Palestine

[3]Department of Medical Laboratory Sciences, Faculty of Health Professions, Al-Quds University, Palestine

[4]Ramallah primary health care, Palestinian Ministry of health

* Corresponding author

E-mail: Kamal.dumaidi@aaup.edu

Abstract:

Background: HAV genotypes and its genetic diversity is rarely investigated in our region as well as worldwide. **Aims**: the aims of the present study were to determine the HAV genotypes and its risk factors and to investigate the genetic diversity of the HAV isolates in the West bank, Palestine. **Study design**: a cohort of 161 clinically and laboratory confirmed HAV (IgM-positive) cases and 170 IgM negative individuals from all the districts of the West Bank, Palestine during the period of 2014-2016 were tested for VP3/VP1 junction of the HAV genome using RT-PCR and sequence analysis. Phylogenetic analysis, genetic diversity and haplotypes analysis were used to characterize the VP3/VP1 sequences.

Results: Overall, all the 34 sequences of the HAV was found to be HAV-IB sub-genotype. The phylogenetic analysis showed four main clusters with cluster III exclusively consisting of 18 Palestinian isolates (18/23-78%) with weak bootstrap values. A high haplotype diversity (Hd) and low nucleotide diversity (π) were observed. Cluster III showed high number of haplotypes (h=8), but low haplotype (gene) diversity (Hd=0.69). A total of 28 active haplotypes with some consisting of more than one sequence were observed using haplotype network analysis. The Palestinian haplotypes are characterized by closely related viral haplotypes with one SNV away from each other which ran parallel to cluster III in the phylogenetic tree. A smaller Palestinian haplotype (4 isolates) was three SNVs away from the major haplotype cluster (n=10) and closer to haplotypes from Iran, Spain, and South Africa. Young age, low level of parent's education, poor hand

washing and drinking of un-treated water was considered the major HAV risk factors in the present study.

Conclusion: HAV-IB subgentype is endemic in Palestine. HAV showed low genetic variation and nucleotide diversity. Furthermore, haplotype network analysis revealed haplotype variation among the Palestinian sequences.

Keywords: Hepatitis A virus, HAV, Hepatitis, Phylogenetic analysis, genetic diversity, Haplotype analysis, Palestine

Introduction:

Hepatitis A virus (HAV) is "a nonenveloped RNA virus belonging to the family Picornaviridae, genus Hepatovirus". HAV is one of the major causes of acute hepatitis worldwide and contributes to substantial morbidity in both developed and developing countries. Based on HAV genome sequences, human HAV have been classified into three genotypes, HAV I, II, III and sub-divided into 6 sub-genotypes (IA, IB, IIA, IIB, IIIA, IIIB) [1].

Worldwide, the incidence rate of the HAV infection is underestimated due to the clinical presentations of this disease, since infection at early childhood is largely passed asymptomatic and/or has mild forms [2]. Recent data showed that the global incidence of HAV is 1.9% with an estimate of 119 million cases of HAV infection [3, 4]. Approximately, 1.4 million new cases of HAV reported each year with up to 22% of the cases being hospitalized [5]. Palestine (West Bank, and Gaza Strip) was classified as an area of very highly endemic area [6]. The Palestinian official figures put the HAV infection incidence rate as high as 9.5-85 per 100,000 during the period between 2000 and 2018 [7]. Actual incidence rate is thought to be higher due to underreporting and asymptomatic cases.

HAV is transmitted mainly by the fecal-oral route and mainly through drinking and consumption of contaminated water and food and to a lesser extend from person to person or via blood transfusion [8-10]. Poor hygiene and sanitation practice reported as the major risk factors for HAV infection, particularly in low and middle-income countries [11].

The World Health Organization (WHO) classified the HAV endemicity based on anti-HAV IgG antibodies as follows: high (≥ 90% IgG seroprevalence by 10 years of age), intermediate (≥ 50% IgG seroprevalence by 15 years of age, < 90% IgG seroprevalence by 10 years of age), and low (≥ 50% IgG seroprevalence by 30 years of age, < 50% IgG seroprevalence by 15 years of age) [12].

The viral infection is characterized by asymptomatic or mild in children at early age of life. On the contrary, adult infections are more frequently occurring with symptoms. In high endemic regions, HAV infection is acquired in early age of childhood and most adult is positive for anti-HAV IgG with a life-long immunity, whereas, in low endemic countries, most adult population is susceptible to infection[3]

A recent review investigated and analyzed data based on anti-HAV seroprevalence in the Middle East and North Africa (MENA) countries reported a gradual shift in the age of HAV infection from early childhood to late child and adult-hood and indicating a shift towards intermediate endemicity in these countries in general [6]. However, based on only one orphan, old study conduct in Gaza between 1995 and 2001, Palestine is still considered a high endemic country. Therefore, no solid conclusion has been made regarding the current endemicity levels of HAV and vaccine recommendations in Palestine [13]

The aims of this study were to determine the HAV genotypes, the risk factors associated with HAV infections and to visualize the genealogical relationship between intraspecific HAV individual genotypes at the population level and exploring the genetic diversity between the HAV sequences.

Materials and methods:

Study sample:

This case-control, cross-sectional study design, comprised of 161 clinically and laboratory confirmed HAV (IgM-positive) cases and 170 apparently healthy controls from all the districts of the West Bank, Palestine during the period of 2014-2016. A special questionnaire was used to collect demographic and medical data including age, sex, residence, education, working status, housing details, toilet facilities (flush vs. pit), drinking water sources (pipe-public net sources vs. collected well or local spring), available used sewage systems (public net system vs. holes), income, educational level of parents, and other socio-economic data. The study was approved by the Palestinian Ministry of Health under the reference number 145/1541/2014. Verbal informed consent was obtained from all patients or their guardians in case of minors. The data was analyzed anonymously.

Serological assays:

All serum samples of both the HAV cases and the healthy subjects were tested for HAV IgM antibodies at the Central Public Health Laboratories, Palestinian Ministry of Health using commercially IgM capture ELISA (Architect, Abbot-USA, ELISA) with a sensitivity and specificity of >99% according to the manufacturer's instructions.

Molecular assays:

Extraction of viral RNA:

The HAV viral genome was extracted from 200µl of serum samples, using a QIAamp Mini Elute Virus spin kit for the viral RNA/DNA extraction (QIAGEN, Germany) according to the manufacturer's instructions, and kept at -20C until testing.

Reverse transcriptase-polymerase chain reaction (RT-PCR):

The HAV genomes of 136 serum samples from IgM- positive patients were amplified using two primers targeting the VP3/VP1 region of the viral genome as described previously by Lee *et al*., (2012) [14]. Briefly, the synthesis of the cDNA and the RT-PCR was carried out in 25µl reaction mixture containing 4µl viral RNA extraction, 10U Reverse transcriptase (AMV), 10pmol of each the forward and the reverse primers: (HAV1; 5' - GCTCCTCTTTATCATGCTATGGAT-3' and rHAV2; 5'-CAGGAAATGTCTCAGGTACTTTC-3') and 12.5µl of PCR Reddy master mix (Thermo Scientific). PCR products (6µl) were loaded onto a 2% agarose gel, electrophoresed, and stained with ethidium bromide for band visualization at an expected length of 244bp using the Gel Doc System 2000 (Bio-Rad Laboratories-Segarate, Milan, Italy). Of the 136 PCR-positive sample, 34 representative PCR amplicons were selected randomly for sequence analysis. The PCR amplicons of the 34 samples were purified and sequenced. The HAV identity search was conducted using GenBank Basic Local Alignment Search Tool (BLAST) http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi.

Genetic diversity analysis

The 23 sequences generated by the study and the 28 sequences retrieved from the Genbank were aligned using the MEGA version X [15]. Maximum likelihood phylogenetic tree with 1000 iterations for bootstrapping was constructed using MEGA version X.

Population nucleotide diversity indices such as nucleotide diversity per site (π), average number of nucleotide differences (k), mean genetic diversity (Hd), genetic differentiation parameters (Fst and Nm) and neutrality tests including Tajima's D and Fu Li's F test were calculated using DnaSP ver. 6.12.03 [16]. The PopArt 1.7 [17] was used to construct a median-joining haplotype network analysis based on country of source of viral genomes to estimate relationship between haplotypes using nexus input files produced by DnaSP version 6.12.03. Haplotype network analysis was colored and edited using a free open-source vector graphics editor called Inkscape 1.0 (www.inkscape.org). The haplotype analysis was double checked by reconstructing the median joining tree using Network 10 (https://www.fluxusengineering.com/) with RDP input file generated by DnaSP ver. 6.12.03 and default parameters of the software including epsilon value of zero and the connection cost method of Röhl for genetic distance calculation [17]. Greedy FHP method for the genetic distance calculation was also used [18].

Risk assessment analysis

The EpiInfo, a free CDC statistical package, was used to analyze data. Odds ratio, 95% confidence interval, and Fisher's exact test were used to assess the risk factors of HAV infection. P-value was considered significant when less than 0.05.

Results

Characteristics of study population

A total of 331 individuals were included in the study with a median age of 15 years and a 1:1 female-to-male ratio. The HAV case group were from all 11 Palestinian districts in the West Bank-Palestine with 35% (56/161) from the Al-Khalil district. All the 34 sequenced samples were shown to be of HAV-IB genotype. Twenty-three sequences were deposited in the gene Bank (Table 1)

Table 1. Demographic data of the sequenced HAV patient samples

Accession No.	Sample code/year of isolation	genotype	Age (yr)	District
MN744241	1B/Pal-2014	HAV-IB	28	Bethlehem
MN744242	11B/Pal-2014	HAV-IB	15	Bethlehem
MN744243	104S/Pal-2016	HAV-IB	6	Salfit
MN744244	41K/Pal-2015	HAV-IB	14	Al-Khalil
MN744245	44Q/Pal-2015	HAV-IB	14	Qalqilia
MN744246	46N/Pal-2015	HAV-IB	25	Nablus
MN744247	73B/Pal-2015	HAV-IB	2	Bethlehem
MN744248	75B/Pal-2015	HAV-IB	16	Bethlehem
MN744249	86K/Pal-2016	HAV-IB	7	Al-Khalil
MN744250	98T/Pal-2016	HAV-IB	9	Tulkarem
MN744251	6B/Pal-2016	HAV-IB	8	Bethlehem
MN744252	8K/Pal-2014	HAV-IB	6	Al-Khalil
MN744253	16K/Pal-2015	HAV-IB	4	Al-Khalil
MN744254	25K/Pal-2016	HAV-IB	7	Al-Khalil
MN744255	26K/Pal-2016	HAV-IB	8	Al-Khalil
MN744256	32K/Pal-2016	HAV-IB	8	Al-Khalil
MN744257	33K/Pal-2016	HAV-IB	13	Al-Khalil
MN744258	44Q/Pal-2016	HAV-IB	14	Qalqilia
MN744259	60R/Pal-2016	HAV-IB	19	Ramallah
MN744260	66K/Pal-2016	HAV-IB	8	Al-Khalil
MN744261	88K/Pal-2016	HAV-IB	12	Al-Khalil
MN744262	89N/Pal-2016	HAV-IB	5	Al-Khalil
MN744263	69T/Pal-2015	HAV-IB	21	Tulkarem

189

190 Phylogenetic analysis

191 Phylogenetic tree of the VP3/ VP1 junction region of HAV genome was conducted using the 23
192 HAV sequences from strains isolated from Palestinian patients along with 28 sequences retrieved
193 from the Genbank. The phylogenetic tree showed four main clusters with cluster III exclusively
194 consisting of 18 Palestinian isolates. Cluster I consisted of 12 strains, of which four were isolated
195 in Palestine (Figure 1). However, the four clusters were weakly supported by bootstrap values (Fig
196 1).

Figure 1. Consensus Maximum Likelihood phylogenetic tree (1000 replicates) based on VP3/VP1 region. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The branch labels in red represent the study sequences, while the black labels represent the sequences retrieved from the gene bank. The different color branches depict different clades. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 51 nucleotide sequences conducted in n MEGA X [15, 19].

197 Genetic differentiation and Diversity

198 Population nucleotide diversity indices and neutrality tests were calculated for the VP3/VP1
199 junction region of the HAV genome, based on phylogenetic clusters (Tables 2 and 3). A high
200 haplotype diversity (Hd) and low nucleotide diversity (π) were observed (Table 2). The total
201 haplotype (gene) diversity (Hd) for the 23 HAV sequences from Palestine and 28 from the
202 Genbank was 0.93± 0.07. At the same time, the total nucleotide diversity per site (π) was 0.01±

0.002, confirming low genetic diversity in the HAV study bulk. The average number of nucleotide differences between any two sequences (k) was 1.7 which is very low. The DnaSP ver. 6.12.03 estimated the total number of haplotypes for the four probable cluster at 24 with highest in cluster I (h=9) (Table 2). Cluster III is composed of 18 isolates, exclusively Palestinian isolates, while one third of cluster I was from Palestine. The Palestinian cluster (III) showed high number of haplotypes (h=8), but lowest haplotype-to-sequence (h:n) ratio (0.4:1), compared to the other clusters and lowest haplotype (gene) diversity (Hd=0.69). Haplotype diversity (H_d), and number of segregating (polymorphic) sites (S) were highest in cluster I which included Palestinian strains (4/12) along with those from other countries (Figure 1); confirming the highest level of genetic diversity between all probable clusters. Nucleotide diversity (π) is equally low in all clusters (0.01± 0.002). The average number of nucleotide differences (k) is also equal in all clusters except cluster I. Tajima's D and Fu-Li's F tests were negative for cluster I, II, and III. The negative values of Tajima's D and Fu-Li's, though not statistically different from neutral expectations supported the low population differentiation. Cluster I and III had relatively the highest negative values for Tajima's D and Fu'Li's F statistics. However, it did not depart significantly from neutrality (P>0.01) (Table 2). Cluster IV recorded a positive Tajima's D value (0.17), but showing no statistically significant departure from neutrality (P>0.01) (Table 2).

Table 2. Haplotype/nucleotide diversity and neutrality tests of the four probable clusters of HAV as calculated for the VP3/VP1 gene

Table 2.

	Haplotype- nucleotide diversity							Neutrality tests	
Cluster	n	h	h:n	Hd±SD	π±SD	K	S	Tajima's D	Fu-Li's F
Cluster-I	12	9	0.8:1	0.90± 0.08	0.01± 0.002	3.3	14	-1.19	-1.59
Cluster -II	7	4	0.6:1	0.71± 0.18	0.01± 0.002	1.1	3	-0.30	-0.51
Cluster-III	18	8	0.4:1	0.69± 0.13	0.01± 0.001	1.1	7	-1.41	-1.72
Cluster-IV	7	4	0.6:1	0.81± 0.10	0.01± 0.002	1.1	3	0.17	-0.03
Total	**44**	**24**	**0.5: 1**	**0.93± 0.07**	**0.03± 0.002**	**3.2**	**23**	**-2.63***	**-5.16****

n**:** Number of sequences, h: Number of Haplotypes, Hd: Haplotype (gene) diversity, π: Nucleotide diversity (per site)[20], K: Average number of nucleotide differences between two randomly chosen sequences from within in the population[21], S: Number of variable/segregating sites.(1 outgroup, 4 have no clear cluster). *: P<0.01. **: P<0.02

Inter-population pairwise genetic distance (Fst) between the four HAV probable populations ranged from 0.46 to 0.65 with Nm value from 0.11 to 0.12 (Table 2) indicating genetic differentiation and minimal migration and gene flow (Nm) between subpopulations. Fst for cluster III which is purely Palestinian compared to clusters I, II and IV were highest (0.52, 0.52 and 0.63, respectively). However, genetic differentiation among subpopulation is generally low especially between clusters I and IV (Gst=0.073). The genetic differentiation between clusters is low as supported by other low genetic differentiation parameters including Gst, Da, and Dxy (Table 3).

Table 3: Gene flow and genetic differentiation indices between the four HAV probable clusters estimated from VP3/VP1gene sequences

Pop 1	Pop 2	Fst	Nm	Kxy	Dxy	Gst	Da
Cluster-I	Cluster-II	0.46	0.12	4.10	0.021	0.098	0.009
Cluster -I	Cluster –III	0.52	0.12	4.75	0.024	0.108	0.013
Cluster-I	Cluster-IV	0.52	0.12	4.61	0.023	0.073	0.011
Cluster-II	Cluster-III	0.52	0.12	2.41	0.012	0.147	0.007
Cluster-II	Cluster-IV	0.53	0.12	2.45	0.012	0.135	0.007
Cluster-III	Cluster-IV	0.63	0.11	3.19	0.016	0.136	0.011

Fst: Wright's F-statistics, pairwise genetic distance[22], Nm: Gene flow and population migration among populations [22, 23], Kxy: Average proportion of nucleotide differences between populations. Dxy: The average number of nucleotide substitutions per site between populations[20], Da: The number of net nucleotide substitutions per site between populations [20], Gst: Genetic differentiation index based on the frequency of haplotypes [24].

Haplotype network analysis:

Figure 2. (a) Median-joining haplotype network of HAV viral haplotypes constructed using PopArt 1.7. (b) Median-joining haplotype network reconstructed using Network 10. The network analysis included 28 multiple viral haplotypes of HAV from 2014-2018. Each circle represents a unique haplotype, color represents country of genome origin, and the size of the circle is proportional to number of viral genomes included. The numbers in bracket in (a) represent the number of single nucleotide variations (SNVs) between haplotypes. The lines on the branches in (b) represent the mutated position with one line per mutation

The median-joining haplotype network constructed by PopArt 1.7 using the 49 taxa produced a total of 28 active haplotypes with some of which consisting of more than one sequence (2-10 sequences) (Figure 2a). Peripheral haplotypes mainly had single nucleotide variation (SNV) from central haplotypes (Figure 4a). The Palestinian haplotypes (red circles) are characterized by closely related viral haplotypes with one SNV away from each other. The Palestinian haplotypes (red) formed a cluster of nodes (n=9) surrounding a major node (haplotype) consisting of 10 identical HAV sequences which ran parallel to cluster III in the phylogenetic tree (Figure 1). A smaller Palestinian HAV haplotype consisting of four sequences was three SNVs away from the major haplotype cluster (n=10) and closer to haplotypes from Iran, Spain, and South Africa which again matches cluster I in the phylogenetic tree.

The reconstruction of median joining haplotype network analysis using Network 10 revealed similar haplotype network analysis profile by both distance calculation methods, the Röhl method and the Greedy HPF method (Figure 2b).

Risk Assessment

The overall, 331 participants (161 cases and 170 health individuals) were analyzed for the HAV risk factors (Table 6). Young age as a demographic variable was found to be significantly associated with HAV infection (OR=7.16, CI: 4.41-11.63, P<0.0001). Level of parents' education was the only socioeconomic risk factor significantly associated with HAV infections (OR=4.72, CI: 2.64-8.41, P < 0.0001). Furthermore, treating drinking water and washing hands before meal as hygiene and behavior risk variables were found to be statistically significant (R=0.11, CI: 0.01-0.965, P< 0.03 and R= 0.24, C:0.15-0.39, P<0.0001) (Table 4).

Table 4. Demographic data, clinical history and the risk factors associated with acquiring HAV infection in the study group; demography, SES, and hygiene and behavior of the HAV cases and controls.

Table 4.

Variable	Category	Case	Control	Total	OR (CI 95%)	P-value
Demography						
Sex	Male	92	85	176	1.32 (0.85-2.03)	0.22
	Female	69	85	154		
	Total	161	170	331		
Age	<14 yrs	111	42	153	7.16 (4.41-11.63)	<0.0001*
	>14 yrs	48	130	178		
	Total	159	172	331		
Socioeconomic status (SES)						
Level of education	≤12 yrs	141	108	249	4.72 (2.64-8.41)	<0.0001*
	>12 yrs	18	65	83		
	Total	159	173	332		
Monthly income	<1450	15	12	27	1.41 (0.64-3.11)	0.43
	>1450	143	161	158		
	Total	158	173	331		
Hygiene and Behavior						
Toilette, location	In-house	154	172	326	0.18 (0.02-155)	0.10
	In yard	5	1	6		
	Total	159	173	332		
Wash hand	Yes	152	166	318	1.14 (0.30-4.34)	1
	No	4	5	9		
	Total	156	171	327		
Livestock at home	Yes	49	61	110	0.76 (0.48-1.18)	0.25
	No	87	75	162		
	Total	136	136	272		
Toilette link	Septic tank	72	77	149	1.03 (0.66-1.59)	0.62
	No	87	96	183		
	Total	159	173	332		
Treat drinking water	Yes	1	8	9	0.11 (0.01-0.965)	0.03**
	No	135	128	263		
	Total	136	136	272		
Swimming	Yes	58	75	133	0.74 (0.47-1.15)	0.21
	No	101	97	198		
	Total	159	172	331		
Wash hands before meal	Always	35	93	128	0.24 (0.15-0.39)	<0.0001**
	Sometimes	124	80	204		
	Total	159	206	332		
Eat leafy vegetables	Yes	102	92	194	1.64 (0.99-2.67)	0.06
	No	34	44	78		
	Total	146	136	272		
Eat in restaurants	Yes	90	110	200	0.72 (0.46-1.13)	0.17
	No	69	61	130		
	Total	159	171	330		

*: statisitcally significant risk factor. **: statistically significant protective factor

Discussion:

To the best of our knowledge, the present study is the first to use genetic diversity indices and haplotype analysis networking to analyze HAV variation from clinical sample. All of the 34 sequenced samples in the present study proved to be of the sub-genotype IB, which is the predominantly circulating genotype in the Mediterranean region from both clinical and environmental samples such as Spain, Jordan, Egypt and Turkey [25-28].

The maximum likelihood phylogenetic tree of the 23 HAV-IB sequences in the present study with other 28 HAV sequences retrieved from the gene bank, identified at least four clusters, but with weak bootstrap values. Although most (78%) of the samples isolated in 2015 and 2016 distinctively clustered in clade III, four Palestinian HAV-IB samples isolated between the years 2014 and 2016 intermixed with other HAV-IB isolates from Turkey, Spain, South African, Iran and Uruguay in clusters I. One isolate clustered uniquely by its own (Figure 1). Similar results had been described in a Bulgarian study which reported the splitting of HAV-IB into several clusters with few cases intermixing between Bulgarian and European isolates while others formed a unique Bulgarian cluster [4]. Furthermore, Wang *et al*., (2013) delineated HAV isolates from the same area and during the same period to have clustered to several closely related lineages with low genetic diversity, suggesting either they possess a fitness advantage in the region, or an endemic transmission of closely related strains circulating in the neighboring regions [29]. The four Palestinian isolates in cluster I-and the lonely clustering isolate may have originated from travelers or food imported from HAV endemic areas, whereas, cluster III that consisted of purely

Palestinian isolates indicates endemic HAV genotypes in Palestine. In addition to the weak bootstrap values, the genetic diversity indices such as total nucleotide diversity per site (π), average number of nucleotide differences between any two sequences (k), and neutrality indices supported the low population nucleotide diversity among HAV sequences.

The negative values of Tajima's D and Fu-Li's F indicate the amount of nucleotide variation observed (π) between HAV isolates is much less than expected (θ) which means low nucleotide variation. HAV clusters I and III with relatively high negative values for Tajima's D and Fu'Li's F statistics and low Hd (0.69) may have been subjected to recent population expansion or selection variation events that reduce genetic diversity such as selective sweep or bottleneck. This low variation between HAV clusters was further supported by the equal values of nucleotide diversity (π) in almost all cluster and the same applied to the average number of nucleotide differences (k). On the other hand, signs of polymorphism and nucleotide variation can be discerned in cluster IV which had positive Tajima's D value (0.17). This cluster consisting of European HAV isolates can be argued to have begun a balancing selection where cluster IV is on the brink of developing into a distinct population (Table 1). Such data had been supported by recent published data from Bulgarian study which showed that the sequence analysis of HAV- IA subgenotypes were either identical or showed very few (1 to 4) nucleotide variations [30].

The fixation indices such as Fst confirmed genetic differentiation between the four HAV probable population with minimal migration among population reducing gene flow (Nm) between the cluster ($Fst > 0.25$ and $Nm<1$).

The total number of haplotypes produced by the haplotype analysis was 28. Indeed, several haplotypes (4-9) were observed in each phylogroup. This proves that despite low genetic diversity indices, haplotypes networking can be a good surrogate for unraveling the diversity among

356 monomorphic population and even within the same phylogenetic group which is due to the fact
357 that haplotype network analysis is based on detecting the single nucleotide variation (SNV) among
358 studied sequences making it a powerful tool to detect genetic and haplotype variation.
359 Interestingly, our study showed that the number of haplotypes in the clusters containing the
360 Palestinian isolates, I and III, were the highest with h= 9 and 8, respectively (Table 2). The high
361 number of haplotypes in cluster III indicate that the Palestinian HAV isolates, despite being from
362 a geographically confined area, are more diverse and heterogeneous than those from Europe
363 (cluster IV). However, despite the high number of haplotypes among our isolates, in comparison
364 to the European cluster, the nucleotides variation is very low (1-2nts) which might be the result of
365 a common origin coming from recent population expansion or selective sweep or genetic
366 bottleneck.
367 Despite the high number of Palestinian haplotypes (red circles), but they are characterized by
368 closely related viral haplotypes with 1-2 SNVs away from each other reflecting how close these
369 haplotypes are. Cluster I is the most heterogeneous as it had the highest number of haplotypes
370 (h=9), the haplotypes ≥2 SVNs apart from each other, and representing isolates from Asia, Europe,
371 South America and Africa.
372 Moreover, the identical median-joining haplotype networks constructed by PopArt 1.7 and
373 Network 10.0 (Figure 3) supported the low-bootstrap value phylogroups created by the maximum
374 likelihood phylogenetic tree using MEGA version X (Figure 2) and running parallel with genetic
375 and haplotype diversity parameters calculated by DnaSP ver. 6.12.03.
376 It is noteworthy, that even though a variety of genetic and haplotype diversity indices and several
377 genetic differentiation parameters were calculated and networks and trees constructed, some

variation may have been missed due to different reasons such as missing nucleotide during sequencing and insufficient number of sequences from study area and from Genbank.

In this study, young age, low level of parent's education significantly increased the odds of HAV infection (OR>1, P<0.05), while the significantly low odds (OR<1, P<0.05) with the increase of hand washing before meal and treating of drinking water indicated the decrease of HAV infection suggesting these two variables are protective factors. On the contrary, Koroglu et al., (2017) reported that water and sanitation were not significant risk factors for HAV infection, whereas, gross domestic product (GDP), gross national income (GNI), and the human development index (HDI) were all highly associated with HAV infection rate in the Middle East and North Africa [6]. In addition, Hayajneh et al., (2015), reported that the incidence rate of HAV infection in Jordan decreased due to increase in level of maternal education, use of processed bottled drinking water and good sanitation practice [31-33]. Our results were in congruence with other studies that reported personal hygiene (hand washing before food preparation, cooking, eating and after defecation), living on crowded campuses, drinking unprocessed water, were the major risk factors of HAV infections [34-37]. In addition, political conflict in Palestine and the occupation of the West Bank and Gaza Strip, Palestine as well as the wars in neighboring Middle East countries usually, disturb the environmental and health condition and commonly increase the spreading of infectious disease outbreaks, including HAV, not only in source countries but also in neighboring regions that host refugees.

In conclusion, this study confirmed that based on VP3/VP1 junction region of the HAV genome; HAV possesses low genetic variation and nucleotide diversity, albeit haplotype network analysis revealed haplotype variation among the Palestinian sequences. Future studies based on multiple targets or full genome sequences are expected to give full picture on the genetic and haplotype

variation. Our study reconfirmed that age and parent's level of education as HAV risk factors, while hand washing and treating drinking water as protective factors.

References:

1. Franco E, Meleleo C, Serino L, Sorbara D, Zaratti L. Hepatitis A: Epidemiology and prevention in developing countries. World J Hepatol. 2012;4(3):68-73. Epub 2012/04/11. doi: 10.4254/wjh.v4.i3.68. PubMed PMID: 22489258; PubMed Central PMCID: PMCPMC3321492.
2. Seo JY, Choi BY, Ki M, Jang HL, Park HS, Son HJ, et al. Risk factors for acute hepatitis A infection in Korea in 2007 and 2009: a case-control study. J Korean Med Sci. 2013;28(6):908-14. Epub 2013/06/19. doi: 10.3346/jkms.2013.28.6.908. PubMed PMID: 23772157; PubMed Central PMCID: PMCPMC3678009.
3. Jacobsen KH, Wiersma ST. Hepatitis A virus seroprevalence by age and world region, 1990 and 2005. Vaccine. 2010;28(41):6653-7. Epub 2010/08/21. doi: 10.1016/j.vaccine.2010.08.037. PubMed PMID: 20723630.
4. Cella E, Golkocheva-Markova EN, Trandeva-Bankova D, Gregori G, Bruni R, Taffon S, et al. The genetic diversity of hepatitis A genotype I in Bulgaria. Medicine (Baltimore). 2018;97(3):e9632. Epub 2018/03/06. doi: 10.1097/MD.0000000000009632. PubMed PMID: 29504993; PubMed Central PMCID: PMCPMC5779762.
5. Fiore AE, Wasley A, BP. B. Advisory Committee on Immunization Practices (ACIP), Prevention of hepatitis A through active or passive immunization: recommendations of the Advisory Committee on Immunization Practices (ACIP). CDC, 2006.
6. Koroglu M, Jacobsen KH, Demiray T, Ozbek A, Erkorkmaz U, Altindis M. Socioeconomic indicators are strong predictors of hepatitis A seroprevalence rates in the Middle East and North Africa. J Infect Public Health. 2017;10(5):513-7. Epub 2017/02/07. doi: 10.1016/j.jiph.2016.09.020. PubMed PMID: 28162965.
7. PalestineMinistryofHealth. Health Annual Report. Ramalla-Palestine: 2018.
8. Bellou M, Kokkinos P, Vantarakis A. Shellfish-borne viral outbreaks: a systematic review. Food Environ Virol. 2013;5(1):13-23. Epub 2013/02/16. doi: 10.1007/s12560-012-9097-6. PubMed PMID: 23412719.
9. Maunula L, Kaupke A, Vasickova P, Soderberg K, Kozyra I, Lazic S, et al. Tracing enteric viruses in the European berry fruit supply chain. Int J Food Microbiol. 2013;167(2):177-85. Epub 2013/10/19. doi: 10.1016/j.ijfoodmicro.2013.09.003. PubMed PMID: 24135674.
10. Hughes JA, Fontaine MJ, Gonzalez CL, Layon AG, Goodnough LT, Galel SA. Case report of a transfusion-associated hepatitis A infection. Transfusion. 2014;54(9):2202-6. Epub 2014/04/03. doi: 10.1111/trf.12648. PubMed PMID: 24689888.
11. Heathcote J, Elewaut A, Fedail S, Gangl A, Hamid S, Shah M, et al. Management of acute viral hepatitis Milwaukee, USA2007 [cited 2020]. Available from: https://www.worldgastroenterology.org/guidelines/global-guidelines/management-of-acute-viral-hepatitis/acute-viral-hepatitis-english
12. WorldHealthOrganization. WHO position paper on hepatitis A vaccines. Weekly epidemiological record. 2012;87(28-29):261-76.
13. Yassin K, Awad R, Tebi A, Queder A, Laaser U. The epidemiology of hepatitis A infection in Palestine: a universal vaccination programme is not yet needed. Epidemiol Infect. 2001;127(2):335-9. Epub 2001/11/06. doi: 10.1017/s0950268801005970. PubMed PMID: 11693511; PubMed Central PMCID: PMCPMC2869753.

14. Lee H, Jeong H, Yun H, Kim K, Kim JH, Yang JM, et al. Genetic analysis of hepatitis A virus strains that induced epidemics in Korea during 2007-2009. Journal of clinical microbiology. 2012;50(4):1252-7. Epub 2012/01/13. doi: 10.1128/jcm.01114-11. PubMed PMID: 22238447; PubMed Central PMCID: PMCPMC3318560.
15. 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):1547-9. Epub 2018/05/04. doi: 10.1093/molbev/msy096. PubMed PMID: 29722887; PubMed Central PMCID: PMCPMC5967553.
16. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol. 2017;34(12):3299-302. Epub 2017/10/14. doi: 10.1093/molbev/msx248. PubMed PMID: 29029172.
17. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16(1):37-48. Epub 1999/05/20. doi: 10.1093/oxfordjournals.molbev.a026036. PubMed PMID: 10331250.
18. Foulds LR, Hendy MD, Penny D. A graph theoretic approach to the development of minimal phylogenetic trees. J Mol Evol. 1979;13(2):127-49. Epub 1979/07/18. doi: 10.1007/BF01732868. PubMed PMID: 480370.
19. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993;10(3):512-26. Epub 1993/05/01. doi: 10.1093/oxfordjournals.molbev.a040023. PubMed PMID: 8336541.
20. Nei M. Molecular Evolutionary Genetics. New York: Columbia University Press; 1987.
21. Tajima F. Evolutionary relationship of DNA sequences in finite populations. Genetics. 1983;105(2):437-60. Epub 1983/10/01. PubMed PMID: 6628982; PubMed Central PMCID: PMCPMC1202167.
22. Wright S. The genetical structure of populations. Ann Eugen. 1951;15(4):323-54. Epub 1951/03/01. doi: 10.1111/j.1469-1809.1949.tb02451.x. PubMed PMID: 24540312.
23. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. Genetics. 1992;132(2):583-9. Epub 1992/10/01. PubMed PMID: 1427045; PubMed Central PMCID: PMCPMC1205159.
24. Nei M. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci U S A. 1973;70(12):3321-3. Epub 1973/12/01. doi: 10.1073/pnas.70.12.3321. PubMed PMID: 4519626; PubMed Central PMCID: PMCPMC427228.
25. Melhem NM, Talhouk R, Rachidi H, Ramia S. Hepatitis A virus in the Middle East and North Africa region: a new challenge. J Viral Hepat. 2014;21(9):605-15. Epub 2014/07/22. doi: 10.1111/jvh.12282. PubMed PMID: 25040644.
26. D'Andrea L, Perez-Rodriguez FJ, de Castellarnau M, Manzanares S, Lite J, Guix S, et al. Hepatitis A virus genotype distribution during a decade of universal vaccination of preadolescents. Int J Mol Sci. 2015;16(4):6842-54. Epub 2015/03/31. doi: 10.3390/ijms16046842. PubMed PMID: 25815599; PubMed Central PMCID: PMCPMC4424991.
27. Cristina J, Costa-Mattioli M. Genetic variability and molecular evolution of hepatitis A virus. Virus Res. 2007;127(2):151-7. Epub 2007/03/03. doi: 10.1016/j.virusres.2007.01.005. PubMed PMID: 17328982.
28. Dinc B, Koyuncu D, Karatayli SC, Berk E, Karatayli E, Parlak M, et al. Molecular characterization of hepatitis A virus isolated from acute infections in Turkey. Turk J Gastroenterol. 2012;23(6):714-9. Epub 2013/06/26. PubMed PMID: 23794310.
29. Wang H, Zheng H, Cao J, Zhou W, Yi Y, Jia Z, et al. Genetic diversity of hepatitis A virus in China: VP3-VP1-2A genes and evidence of quasispecies distribution in the isolates. PLoS One. 2013;8(9):e74752. Epub 2013/09/27. doi: 10.1371/journal.pone.0074752. PubMed PMID: 24069343; PubMed Central PMCID: PMCPMC3775754.

30. Bruni R, Taffon S, Equestre M, Cella E, Lo Presti A, Costantino A, et al. Hepatitis a virus genotypes and strains from an endemic area of Europe, Bulgaria 2012-2014. BMC Infect Dis. 2017;17(1):497. Epub 2017/07/15. doi: 10.1186/s12879-017-2596-1. PubMed PMID: 28705178; PubMed Central PMCID: PMCPMC5513050.

31. Hayajneh WA, Balbeesi A, Faouri S. Hepatitis A virus age-specific sero-prevalence and risk factors among Jordanian children. Journal of medical virology. 2015;87(4):569-74. Epub 2015/02/05. doi: 10.1002/jmv.24137. PubMed PMID: 25648328.

32. Toukan AU, Sharaiha ZK, Abu-el-Rob OA, Hmoud MK, Dahbour SS, Abu-Hassan H, et al. The seroepidemiology of hepatitis A virus infection in Jordan. Trop Gastroenterol. 1988;9(2):76-9. Epub 1988/04/01. PubMed PMID: 3262944.

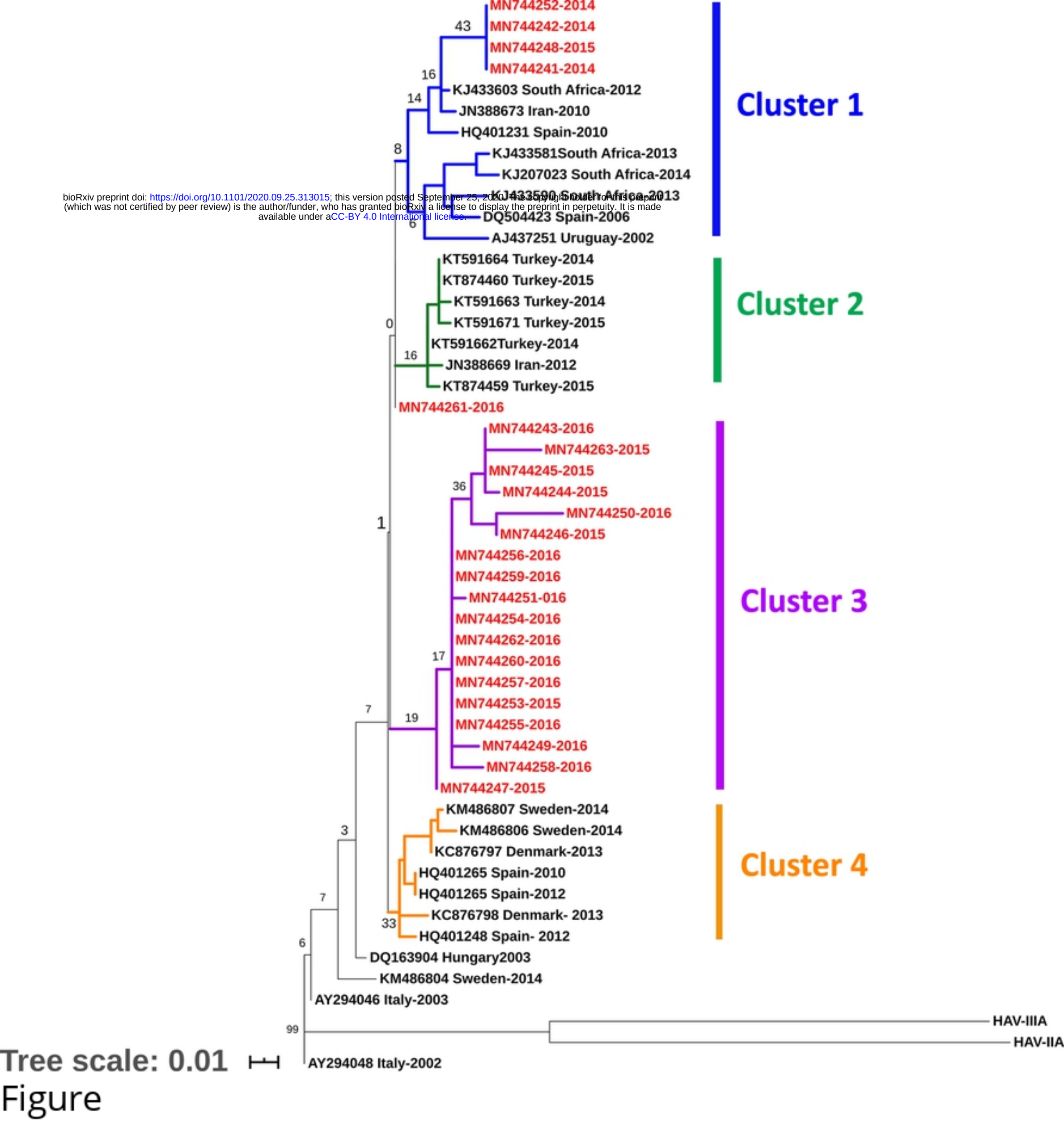
33. Battikhi MN, Battikhi EG. The seroepidemiology of Hepatitis A virus in Amman, Jordan. The new microbiologica. 2004;27(3):215-20. Epub 2004/10/06. PubMed PMID: 15460523.

34. Juniastuti, Wahyuddin D, Nihayatussa'adah, Amin M, Yamani LN, Utsumi T, et al. Analysis of genetic and serology of hepatitis A virus infection during and after outbreak in two junior high schools in Surabaya, Indonesia. Journal of medical virology. 2019;91(6):1048-55. Epub 2019/01/20. doi: 10.1002/jmv.25403. PubMed PMID: 30659645.

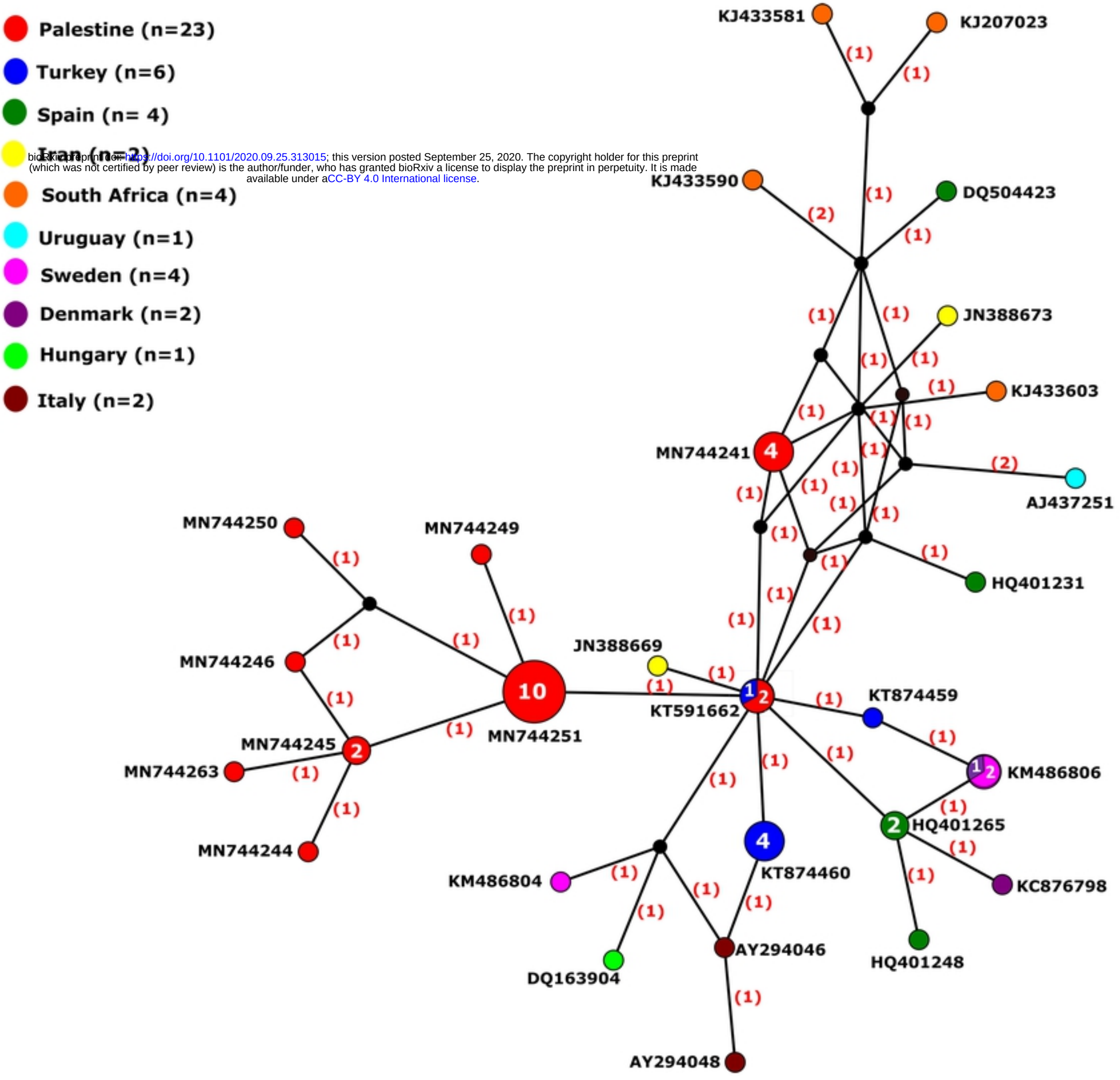
35. Kotwal A, Singh H, Verma AK, Gupta RM, Jain S, Sinha S, et al. A study of Hepatitis A and E virus seropositivity profile amongst young healthy adults in India. Med J Armed Forces India. 2014;70(3):225-9. Epub 2014/11/08. doi: 10.1016/j.mjafi.2014.06.016. PubMed PMID: 25378774; PubMed Central PMCID: PMCPMC4213908.

36. Pinheiro RS, Araujo LA, Caetano KA, Matos MA, Carneiro MA, Teles SA. Intermediate Endemicity of Hepatitis a Virus Infection in Rural Settlement Projects of Southwest Goias, Brazil. Arq Gastroenterol. 2015;52(3):200-3. Epub 2015/10/22. doi: 10.1590/S0004-28032015000300009. PubMed PMID: 26486287.

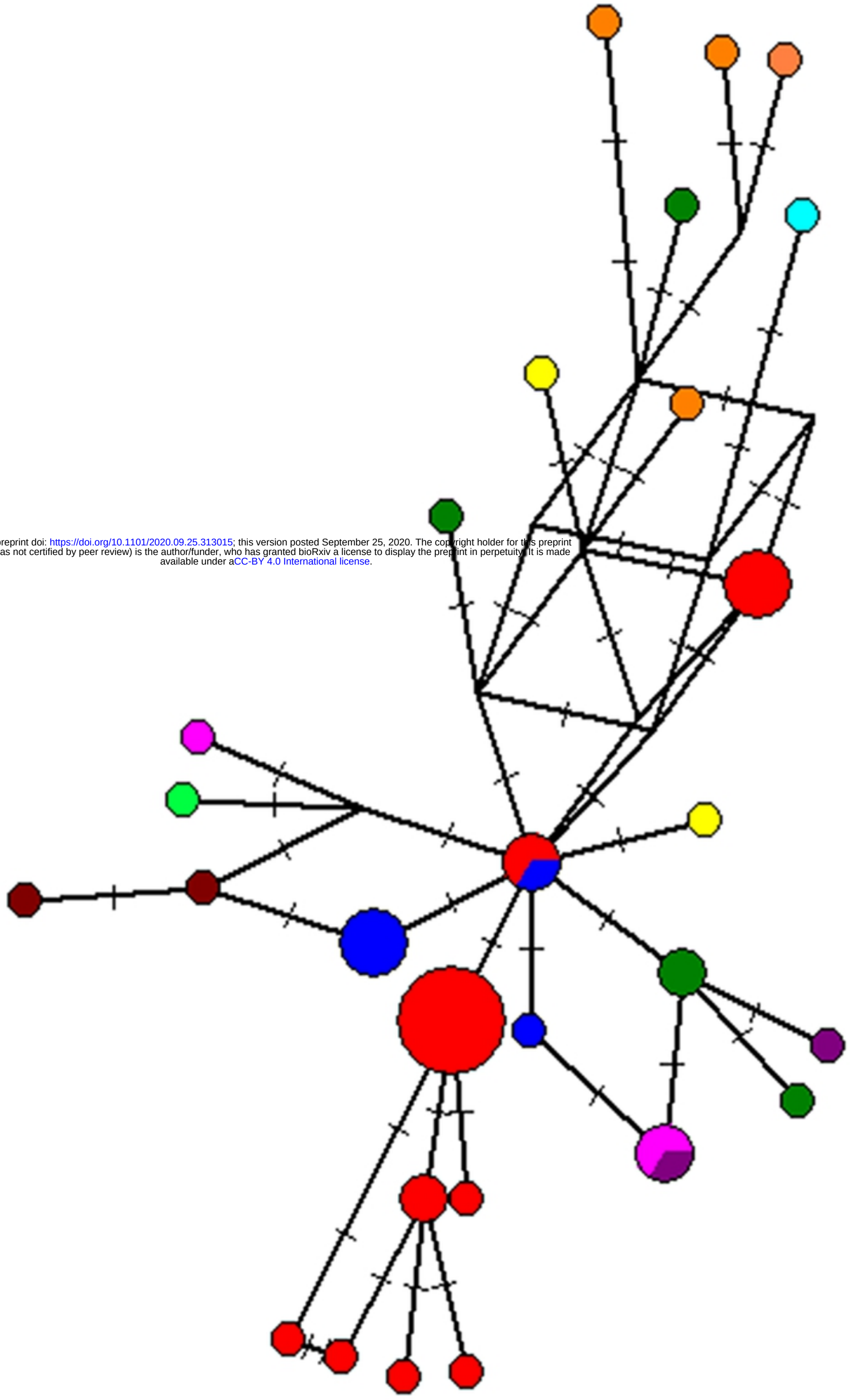
37. de Almeida LM, Amaku M, Azevedo RS, Cairncross S, Massad E. The intensity of transmission of hepatitis A and heterogeneities in socio-environmental risk factors in Rio de Janeiro, Brazil. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2002;96(6):605-10. Epub 2003/03/11. doi: 10.1016/s0035-9203(02)90325-1. PubMed PMID: 12625132.



Figure



Figure



Figure