Low Seroprevalence of Torque Teno Virus in HCV positive patients and phylogenetic analysis from Pakistani isolates

Amen, N.[1], Maria, M.[1], Azam, M.[1], Aziz, A.[1], Qamar, R.[1,2*] and Bostan, N.[1*]
[1]COMSATS Institute of Information Technology, Islamabad, Pakistan
[2]Al-Nafees Medical College and Hospital, Isra University, Islamabad Pakistan
*Corresponding author e-mail: raheel_qamar@comsats.edu.pk, nazishbostan@comsats.edu.pk
Received 10 November 2016; received in revised form 2 October 2017; accepted 4 October 2017

Abstract. The torque teno virus (TTV) has a heterogeneous genome and its role in hepatitis C (HCV) infection is still controversial, therefore the purpose of the present study was to determine if there is any association between Hepatitis C and TTV co-infection and to determine the phylogenetic relationship between existing types in the Pakistani population. A total of 500 individuals (250 HCV positive patients and 250 healthy controls) were selected. DNA was extracted from serum samples and polymerase chain reaction (PCR) of the open reading frame 1 (ORF1) region of TTV was performed. Out of 500 samples 9 HCV positive index cases (3.6%) and 8 healthy samples (3.2%) were found to be positive for TTV. A comparison was made between TTV sequences reported from all over the world with the ones obtained in the present study by sequencing of TTV positive samples followed by phylogenetic analysis using maximum parsimony (MP) method. Our results indicated that the virus was undergoing divergent evolution as very high sequence diversity was found in the ORF1 gene. The study also shows that association between HCV and TTV was not found. Because the virus was found to be affecting both healthy individuals and HCV infected population with almost same frequency. Therefore a thorough screening of TTV virus at the population level is required in order to draw a comprehensive inference.

INTRODUCTION

Torque teno virus (TTV) is one of the newly identified members of genus *Anellovirus*, which was discovered in 1997. The genome size of TTV is ~3.6-3.8 kb, which is a small, circular and single stranded DNA that makes it a unique type of viral genome reported in human beings (Nishizawa *et al.*, 1997). The heterogeneity in the genome of TTV makes it a very diverse group of related viruses with at least 35 reported genotypes. The untranslated regions (UTRs) of the viral genome are considered relatively more conserved as compared to its three overlapping open reading frames (ORFs) (Erker *et al.*, 1999) encoding three mRNAs, however, the functions and locations of possible protein products are still vague (Karen, 2015; Mankotia & Irshad, 2014).

Since its discovery from a hepatitis patient of unknown etiology, it has been a focus of research with reference to its association with other hepatic viruses like hepatitis B virus (HBV), hepatitis C virus (HCV) and SEN virus causing hepatocellular carcinomas and other liver related ailments. Besides its co-infection with hepatic viruses it is considered to be a causative agent of many infections like acute respiratory diseases, cancer, AIDS and endophthalmitis (Maggi *et al.*, 2003a; Maggi *et al.*, 2003b). However, the findings are vague (Maximova *et al.*, 2015). The prevalence TTV varies from country to country and is found to be more common in healthy populations worldwide. Previous studies found a high seroprevalance of this virus both in healthy individuals including blood donors and also in liver disease index cases (Mazzola *et al.*, 2015). It

has been almost 18 years since its discovery but the molecular pathways that TTV follows in the host cell and its pathogenicity is still unclear. Cell culture propagation difficulty, omnipresence of TTV, its genotype variability and inability to design a single set of primer for the amplification of its genome makes its study challenging (Vasilyev *et al.*, 2009). Till now there is only one report on the presence of TTV in the Pakistani population, which is based on HBV, HCV and healthy population controls but in a small cohort (Hussain *et al.*, 2012).

The present study compiles the first indigenous report to comprehensively illustrate the incidence of TTV in uninfected and hepatitis C Virus infected population from Pakistan. The ORF1 region of TTV was also sequenced, annotated through *in silico* analysis and the phylogenetic relationship with other viral isolates was studied.

METHODOLOGY

Sample Collection

The study was approved by the "Ethics Review Board" of the Department of Biosciences, COMSATS Institute of Information Technology, Islamabad, Pakistan. A written informed consent was taken from each participant prior to sample collection. In the present study, samples of 250 HCV positive patients (from out-door patient department (OPD) of Shifa International Hospital) and 250 population matched healthy controls were selected. Blood (3ml) was drawn from all study participants in lavender top vaccutainers tubes with K_2EDTA as anticoagulant to prevent blood clotting (BD Vacutainer®). A proforma was designed to determine the important aspects of demography and disease history of the sampled individuals. The cohort was divided into five age groups i.e., (15-20), (21-30), (31-40), (41-50), (51-60). For each age group 50 individuals participated in the study on volunteer basis. In HCV positive patients the female patients (136) were slightly more than male patients (114). On the other hand in the case of control group there were more males (190) than females (60). In the general public

it is a common phenomenon that females avoid participating in a study where blood sampling is involved; therefore the number of females in the control group was less than the male participants.

Inclusion and Exclusion Criteria

Both the HCV cases and control individuals were above 15 years of age, as the target was to screen the adult population, therefore children and adolescents were not selected. The control individuals were free of any past or present diseases like cancer or viral infections. For HCV positive sample collection, only naive diagnosed cases were selected for this study as interferon treatment can alter the results.

Confirmation of HCV infection

HCV infection was detected through serological testing using QuickTiter™ Hepatitis C Core Antigen (HCV_cA_g) ELISA Kit for the detection of anti HCV antibodies as per manufacturer's protocol (Cell Bio Labs, Inc, San Diego, CA 92126). Further confirmation of HCV infection was carried out with HCV RNA testing using a ready-to-use system for the isolation and detection of HCV from plasma/serum using end-point RT PCR Kit following the manufacturer's protocol (Norgen Biotech Corp, Canada).

DNA Extraction and PCR analysis

DNA was extracted from the blood using QIAamp Viral DNA Mini Kit 51306 (Qiagen Netherlands). The extracted TTV DNA was subjected to PCR using a set of 3 semi nested primers as described previously by (Prescott *et al.*, 1999) and were designed against ORF1 (Abbass *et al.*, 2011). The PCR samples were separated on 2% agarose gel using TBE buffer. 100 bp DNA ladder, positive control and negative control were loaded in each set of gel (Figure 1). The gel was visualized under a UV trans-illuminator (BioDoc Analyzer, Biometra (BDA BOX2).

DNA Purification and Sequencing

Samples were sequenced using first ORF1F2 primer and then using ORF1R primer from ORF1 of the TTV genome using Jet Quick PCR Product purification spin Kit by Genomed.
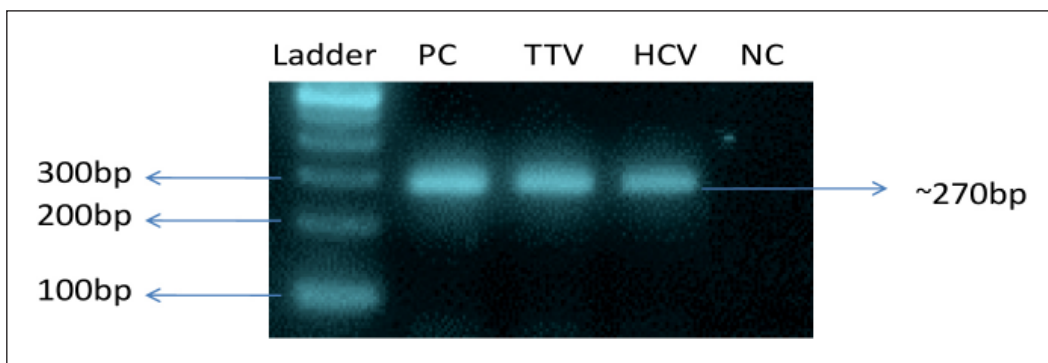
Figure 1. Results of agarose gel electrophoresis from ORF1 region showing the positive samples of control and HCV (NC=negative control; PC=Positive control). 2% Agarose gel in 1× TAE buffer. 100bp DNA ladder promega G2101.

The sequenced samples were purified and analyzed using software BioEdit version 7.0.5 (Hall *et al.*, 2011). Basic local alignment search tool (BLAST) was applied to the sequenced samples to check their homologies. Fasta format was used for the sequenced samples, which was then used as an input of Clustal W and SPA (Short Peptide Assembler) software (Yang & Yooseph, 2013) for alignment (Neumann *et al.*, 2014). For all sequenced samples both the sequences were aligned, overlapped using Clustal W software and were further processed for their submission to GenBank and allotment of accession numbers (Larkin *et al.*, 2007). Accession numbers assigned by GenBank were JX885493-JX885509.

Phylogenetic Analysis
Statistical selection pairing by applying Tajima's test (Tajima, 1989), nucleotide composition and Maximum Parsimony (MP) (Tamura *et al.*, 2011) methods were used for phylogenetic reconstruction and the *p-distance* model was used for distance analysis (Thomas, 2001). Base statistical robustness was performed by using Maximum parsimonious trees in original with *p-distances* and with 500 repeat boot strap analysis. In order to authenticate the results bootstrapping was conducted, and the significance cut off values were 75% and above and the whole process was analyzed by MEGA 5.10 Beta # 4 software (Tamura *et al.*, 2011). Disparity Index test, which is more powerful than chi square test was applied to all 17 sequences as observed difference in pattern of substitution in a pair of sequences is measured by it. The reason was to test the base frequencies equality between all TTV sequences (Tamura *et al.*, 2011; Kumar & Gadagkar, 2001). The purpose of constructing Phylogenetic trees and dendrograms was to find out the relatedness and evolutionary relationship among the sequences. Both host and viral factors actively play a role in tree topology determination. To determine the relationship in terms of sequence similarities between the sequenced samples dendrogram analysis was done (Figure 2). The height of vertical lines in the dendrogram is the indicator of degree of differences in the isolates of TTV. Varying lengths of dendrogram branches indicate that the TTV isolates are dynamic and are different from each other.

Comparative Phylogenetic Analysis of Studied TTV Sequences with other Sequences from all over the World
For the comparison of the studied TTV sequences with the reference accessions already reported from Pakistan, an exhaustive search was conducted of GenBank (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and European Molecular Biology Laboratory (EMBL) (http://www.ensembl.org/index.html) databases for ORF1 sequence retrieval. However no TTV ORF1 sequence was found for Pakistan in GenBank. Therefore the available sequences reported by different groups including Brazil, USA, Iran, China,
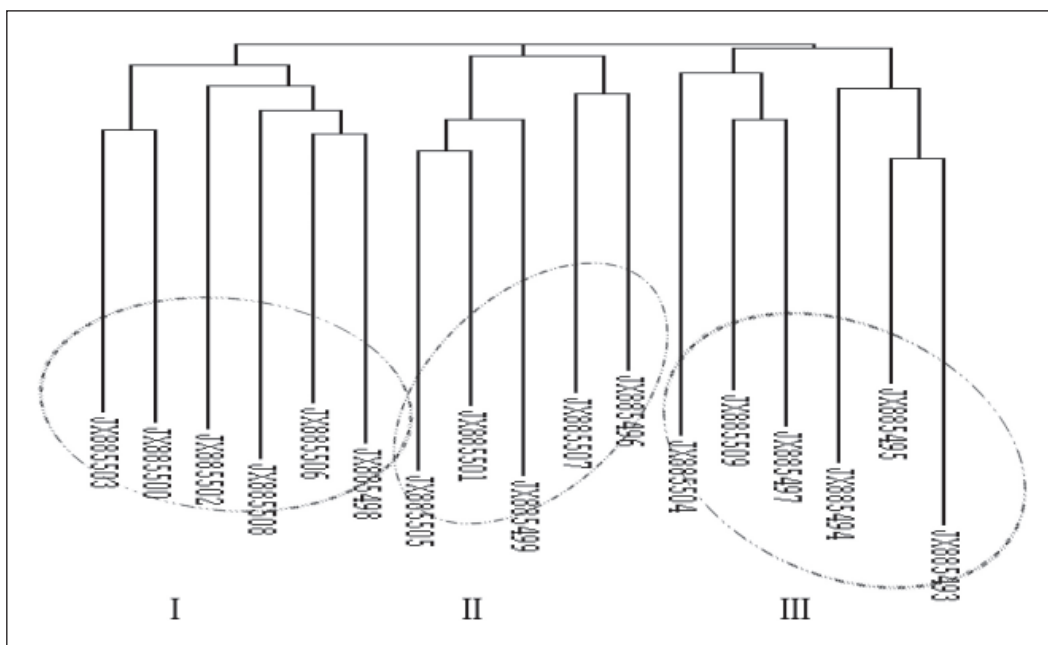
Figure 2. Dandrogram tree of TTV sequences from present study generated by ClustalW method with 1000 boot strap repeats, from Pakistan.

Thailand, Poland, Germany, Italy and Japan were indiscriminately retrieved from NCBI for phylogenetic analysis. As the viral genotypes vary from country to country, the purpose of this work was to compare the similarities and differences between the studied TTV strains and to find out their evolutionary relationship with the previously reported accessions (Figures 3 & 4). The accessions were AY147063 Brs, AF067978 Ger, AY032892 Itl, AY032886 Itl, AF067979 Ger, AJ309721 Pol, AF151683 Chn, AF416141 Chn, AY032895 Itl, AF067977 Ger, AF072868 Thlnd, AF151532 Chn, AB017767 Jap, AF082743 Chn, AF055897 Chn, AF073795 Thlnd, AY032887 Itl, AY147064 Brs, FJ597965 Chn, AJ300743 Pol, AY032893 Itl, AY032889 Itl, AY032890 Itl, AY032896 Itl, AJ309724 Pol, AF395926 Grec, AY032888 Itl, DQ195380 USA, AF395922 Grec, GQ337058 Irn, GQ179964 Irn, GQ179967 Irn, AY147052 Brs, AY147050 Brs, AF416142 Chn, AF250216 Thlnd, AF250217 Thlnd, AF250215 Thlnd, AF250213 Thlnd, AF250218 Thlnd and AF250214 Thlnd.

RESULTS

TTV prevalence in studied cases

The prevalence of TTV in the HCV positive samples and in the controls determined by PCR using ORF1 region was 3.6% (5 males and 4 females, n=250) and 3.2% (5 males and 3 females, n=250), respectively. The ages of HCV positive individuals suffering from TTV was one male patient in the 15-20 years age group, one female in 21-30 year age group, one male and female each in age group 31-40, two males and one female in age group 41-50 and 1 male and 1 female in 51 year and above age group. Similarly there were one male and one female each in the 21-30 years age group, two males and one female in 31-40 year age group, one male and one female in 41-50 year age group and one male in $\geq$51 age group. Overall it was observed that most affected age groups in control as well as HCV positive cases were 31-40 (5 cases) and 41-50 year age group (5 cases).
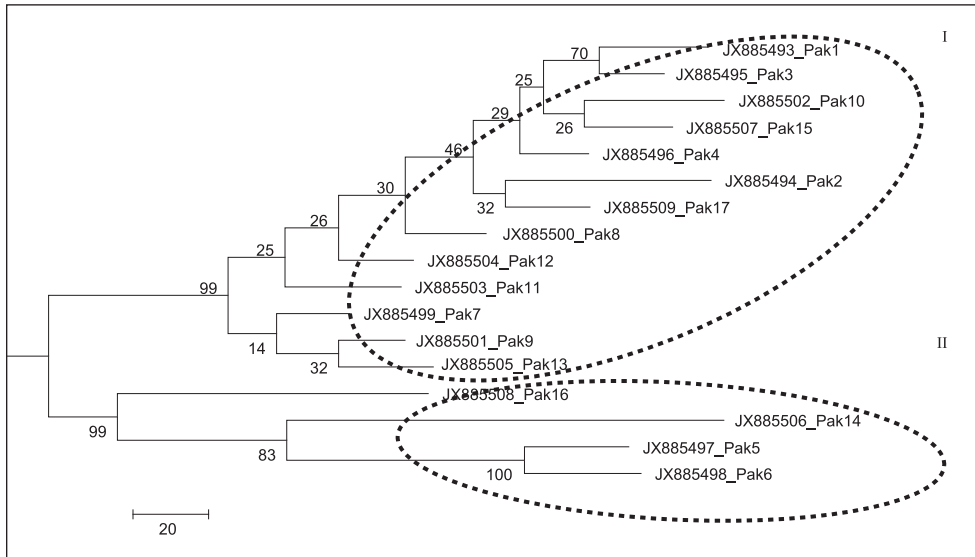
Figure 3. Phylogenetic tree of TTV sequences from present study drawn by Maximum parsimony method with 500 boot strap repeats, from Pakistan.
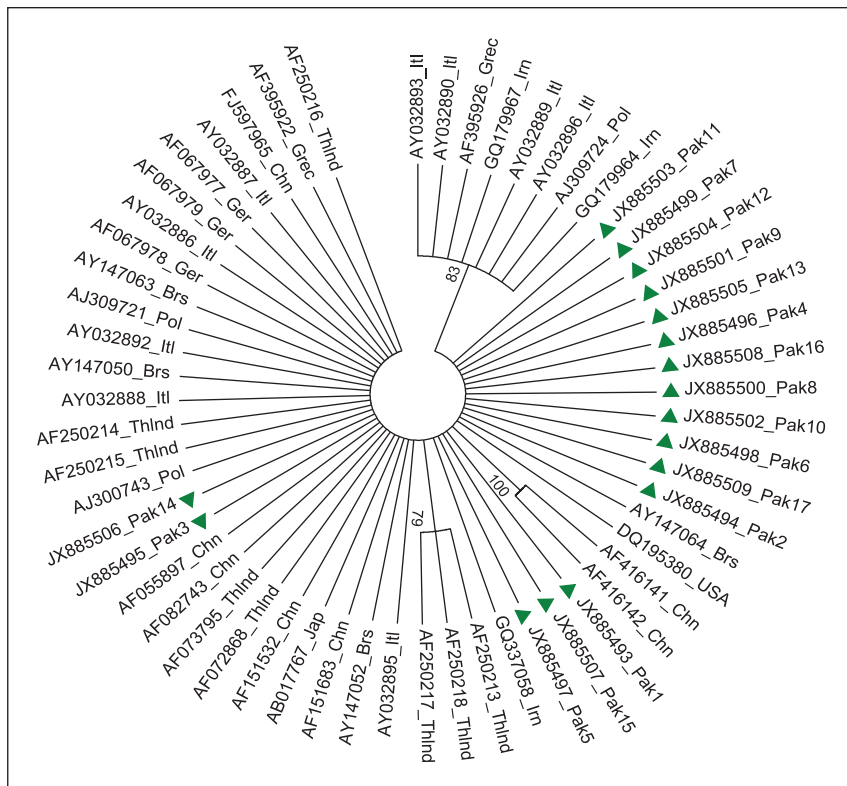

Figure 4. Phylogenetic tree constructed by maximum parsimony method with 500 bootstrap repeats, indicating a comparative tree of accessions retrieved from world over with studied samples.

Sample IDs of the present study are represented with black triangles.

Sequencing of TTV positive cases and BLAST analysis

Nucleotide composition of TTV isolates

A stretch of ~271 nucleotide was amplified by PCR and was sequenced. On an average Thymine/Uracil was 21.9 nucleotides, Cytosine was 22.2 nucleotides, Adenine was 36.8 nucleotides and Guanine was 19.1 nucleotides. In total the percentage of adenine was maximum, imparting stability to the DNA.

To verify the obtained sequences of 17 TTV samples Blast analysis (NCBI http://blast.ncbi./nlm.nih./gov/Blast.cgi/ and EMBL http://www.ebi./ ac.uk/Tools// blast2// nucleotide.html) was performed to determine the percentage sequence similarities with already reported sequences globally. BLAST analysis showed a high sequence match scores between the analyzed sequences and already reported sequences of TTV ORF1 region. After verification the sequences were submitted to GenBank for allotment of accession numbers.

Phylogram and dendrograms of studied TTV sequences

It is evident from both phylogram and dedrogram (2, 3 and 4) that the genome of the virus is an active example of divergent evolution. All texa have evolved over the course of time and the unrooted tree indicates that the variants do not have a common ancestor. The original MP tree (Figure 3) shows divergent evolution with significant homologies and relatedness between sample IDs JX885508, JX885506, JX885497 and JX885498. These entire four sister texa have evolved from a single common ancestor at secondary node level and formed a cluster while all other texa remained independent. The tree topologies were observed for these samples in the tree, with 500 bootstrap repeats (with 99%, 83% and 100% boot strap support at primary secondary and tertiary nodes). All other accessions have evolved independently during viral evolution and their branch topologies show stability with very low mutation rates.

Maximum parsimonious tree with 500 boot strap revealed that the TTV accessions amplified from Pakistan had a common ancestor and were in the phase of active evolution. There were two clusters, in cluster one actively evolving texa were placed. All texa were significantly associated at the primary node level with 99% boot strap support. While at the secondary, tertiary and quaternary level the homologies were not significant.

Rooted tree pointed out that there was a most recent common ancestor (MRCA) for all the descendants. The numbers on the branches on the tree show the number of times the partition occurred of the species into the two sets, which are divided by that branch, occurred among trees, out of 100.00 trees.

In silico analysis of studied accessions with sequences retrieved from world over

The MP tree dragged for the studied samples with reference accessions from world over, mounted seven main clusters where all representative samples of the present study showed independent evolution but not from a MRCA, as the tree is unrooted. Studied samples aligned with their previously established sister subgroups and showed divergent evolution as no additional variation in tree topologies was observed, in terms of, homology between studied samples and reference accessions of the world. All studied homologies were non-significant in terms of *p-distance* values and boot strap support. Significant homologies were found in accessions AF250217, AF250213 and AF250218 from Thailand with 79% bootstrap support. Samples from Iran, Italy, Poland and Greece evolved from a common ancestor at the same time and their *p-distance* values were significant. Two of the Chinese accessions AF416141 and AF416142 were very near sister subgroups (with 100% boot strap support). The evolutionary pattern was the same for the accessions retrieved from all over the world and the studied samples. These accessions depict that the behavior of virus remained stable during the course of evolution.

Tajima's Test of Neutrality

The analysis involved 17 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 270 positions in the final data set. Evolutionary analyses were conducted in MEGA5.10 BETA#4 (Tamura *et al.*, 2011). The positive value of D 0oo(1.912003) in Tajima's Test of Neutrality in Table 1 indicates that polymorphism between sequences of ORF1 Region of TTV was probably maintained by the process of balancing selection and the process remains continuous by an evolutionary method of gene birth and gene death. Polymorphism by host selection was soughed by such combinations.

Test of the Homogeneity of Substitution Patterns between Sequences /Disparity Index Test

Disparity index test of substitution pattern homogeneity was performed for the TTV sequences. The number of Monte Carlo Replications was set at 500. Nucleotide substitution was used, and missing data were treated with complete deletion, total number of sites were 270. The test was performed using MEGA 5.10 BETA# 4 software. The results are depicted in the form of a table (Table 2). Where the computed probability was in black colored letters. The confidence interval value to compute probability was 95% [$\leq$0.05 % significance level]. Values of disparity index were in red color.

It is one of the methods that take, into account the nucleotide frequency bias for e correction of multiple substitutions (it has to be assumed that the pattern of nucleotide substitution has remained uniform during the evolutionary history of the considered sequences). In a data set of 17 TTV sequences it is important to note whether or not all the sequences have evolved with the same substitution pattern. At times the evolutionary pattern is heterogeneous instead of being homogenous. Such sequences provide valuable information in the shift of mutational patterns and selection pressure. For the rejection of null hypothesis of homogenous substitution pattern p-value, compositional distance and normalized values of disparity index play a vital role. Pairwise distance between JX885496 and JX885499 provided the p-value of 0.02 and JX885505 and JX885507 provided the p-value of 0.006 (highlighted in yellow as shown in Table 2). Disparity index test provided valuable information in the shift of mutational pattern.

DISCUSSION

From the time of discovery of TTV, its pathogenicity has been linked with hepatic disorders. TTV has a shared mode of spread with HBV, HCV and other blood borne viruses. Moreover some TTV genotypes may be more infectious than others, and may have a role in causing hepatic problems (Nishizawa *et al.*, 1997). On the other hand, some studies rank TTV as a commensal virus (Bernardin *et al.*, 2010; Fuels & Gut, 2012; Hino & Miyata, 2007). The co infection of TTV with HCV has been widely studied; however, its role as 'helper' or 'causative' agent in hepatic diseases is still uncertain (Tokita *et al.*, 2002). Several reports from different parts of the world like Myanmar, Saudi Arabia, Republic of Guinea-Bissau and United States of

Table 1. Tajima's Test of neutrality applied to compute the level of divergence in TTV variants

m	S	p_s	θ	π	D
17	258	0.955556	0.282648	0.408170	1.912003

The Tajima test was calculated using MEGA 5.10 BETA #4 software. All gaps were eliminated from the data group (complete deletion option). m = Number of sites, S = number of segregation sites, P_s = S/m, ϑ = ps/a1, π = nucleotide diversity. D is the statistical test result.

Table 2. Disparity Index test. Numerics with bold font represent the significant associations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
JX885493_Pak1		0.000	0.000	0.000	0.000	0.000	0.141	0.000	0.000	0.107	0.000	0.000	0.111	0.000	0.000	0.000	0.000
JX885494_Pak2	1.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
JX885495_Pak3	1.000	1.000		0.000	0.000	0.000	0.256	0.174	0.000	0.252	0.200	0.000	0.419	0.000	0.000	0.000	0.078
JX885496_Pak4	1.000	1.000	1.000		0.000	0.000	0.419	0.144	0.000	0.333	0.074	0.000	0.333	0.000	0.000	0.000	0.059
JX885497_Pak5	1.000	1.000	1.000	1.000		0.122	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
JX885498_Pak6	1.000	1.000	1.000	1.000	0.160		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
JX885499_Pak7	0.224	1.000	0.134	0.020	1.000	1.000		0.000	0.144	0.000	0.000	0.089	0.063	0.000	0.348	0.000	0.000
JX885500_Pak8	1.000	1.000	0.192	0.170	1.000	1.000	0.150		0.000	0.000	0.000	0.000	0.000	0.000	0.233	0.000	0.033
JX885501_Pak9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		0.000	0.000	0.000	0.070	0.000	0.000	0.000	0.000
JX885502_Pak10	0.254	1.000	0.098	0.054	1.000	1.000	1.000	1.000	1.000		0.000	0.000	0.000	0.000	0.330	0.000	0.000
JX885503_Pak11	1.000	1.000	0.172	0.300	1.000	1.000	1.000	1.000	1.000	1.000		0.000	0.000	0.000	0.296	0.000	0.000
JX885504_Pak12	1.000	1.000	1.000	1.000	1.000	1.000	0.246	1.000	1.000	1.000	1.000		0.000	0.000	0.000	0.000	0.000
JX885505_Pak13	0.258	1.000	0.282	0.064	1.000	1.000	0.280	1.000	0.264	1.000	1.000	1.000		0.000	0.556	0.000	0.000
JX885506_Pak14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		0.000	0.000	0.000
JX885507_Pak15	1.000	1.000	1.000	1.000	1.000	1.000	0.050	0.126	1.000	0.050	0.084	1.000	0.006	1.000		0.000	0.078
JX885508_Pak16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		0.000
JX885509_Pak17	1.000	1.000	0.226	0.290	1.000	1.000	1.000	0.310	1.000	1.000	1.000	1.000	1.000	1.000	0.238	1.000	

America (Armstrong *et al.*, 2006; Plamondon *et al.*, 2007) are in agreement with this study where 61%, 53% and 51% respectively of the study participants were females and only 39%, 47% and 49% respectively were males. The reason may be the natural high prevalence in females as they are more exposed to blood borne infections during their visits to hospitals in pregnancy and birth processes.

One of the major reasons for a low occurrence of TTV in the study participants might be the low sero-positivity rate of TTV DNA. There is a possibility of low prevalence of this virus in the Pakistani population like some other countries of the world like USA, etc (Irshad *et al.*, 2006). The seropositivity of TTV was 3.6% in HCV infected individuals and was 3.2% in control group, while its seroprevalence in USA, and in European countries (Hino & Miyata, 2007) is 1%. On the other hand in many other parts of the world its prevalence is very high like in Japan it ranges between 39-68%, in Hungry 18.5%, in Brazil 62% and in Germany it is 14% (Mauss *et al.*, 2017).

Reported in many studies, the primer sequences and the PCR conditions, used in experimentation, greatly vary in amplifying variable number of genotypes and geno-groups (Ali *et al.*, 2004; Blazsek *et al.*, 2008; De Villiers *et al.*, 2009). Another possible reason for the low prevalence can be the sample size, normally large samples are more predictive and accurate conclusions can be drawn from them. Yet another reason for such a low positivity could be low titter of TTV DNA in serum and natural recovery from TTV DNA. It is suggested by some scientists that by using different amounts of serum samples, TTV DNA positivity changes, i.e., with 10 µl samples the positivity rate was 1.9% but with 100 µl sample the positivity rate obtained was 4.0% (Simmonds *et al.*, 1998). Hence, the amount of serum samples also affects the positivity rate.

From various studies it has been observed that the TTV viremia is not persistent throughout life (Bendinelli *et al.*, 2001; Maximova *et al.*, 2015). Natural recovery from TTV viremia is a known phenomenon, and as PCR assays are unable to check the past TTV infection history, so a serological test should be performed to check the past infections. This would be of great epidemiological value in investigation of total TTV exposure (Běland *et al.*, 2014; Görzer *et al.*, 2015). The results of the current study showed that no correlation between TTV and HCV exists. The correlation between HCV and TTV is very controversial, some researchers have shown that there is a co-relation between TTV and HCV (Iver *et al.*, 2005), while others could not find any co-relation (Jones *et al.*, 2005; Kheradpezhouh *et al.*, 2007; Kim *et al.*, 2000). The present study found only one co-infected patient out of 182 HCV positive patients, which means 0.5% positivity rate in co-infected patients. Thus it can be said that with the use of ORF 1 primers, no co-relation between TTV and HCV was observed.

There was an option of applying two methods for the construction of phylogenetic trees i.e., either Neighbor Joining method (NJ) or Maximum Parsimony method (MP). Both have their own merits and demerits. In the present study after extensive literature search the MP method was used for construction of the Phylogenetic trees. In a previous study on the evolution of TTV in Iran the authors compared these two methods for tree construction and they found the MP method to be more accurate than the NJ method. The results suggested that classification of TTV by maximum parsimony (MP) method might be more precise than by neighbor-joining (NJ) method (Kenar Koohi *et al.*, 2012). The results of cladistics applied to TTV sequences from Pakistan revealed that all 17 characterized strains were of human TTV. It also showed that virus is evolving independently and suggests the viral evolution to be divergent. Similar results were reported by (Hino & Miyata, 2007) who observed one of the reason for diversity in TTV to be frequent recombination, other factors for this diversity are still unexplained. It is still unclear how this virus diversifies and acquires variation. According to estimates more than 90% of the world population is exposed to this virus but with no apparent disease associated with it, which leads to a suspicion of it being a

commensal human virus. Several other reports indicate that it is evident that in TTV extensive genetic variability exists (Diniz-Mendes *et al.*, 2008; Okamoto *et al.*, 1999; Peng *et al.*, 2002; Peng *et al.*, 2015). In the present study, classification of isolates into genotypes and genogroups was not possible because of limited nucleotide sequence information obtained, but one evident finding was the large diversity of TTV isolates circulating in the blood stream of healthy and HCV infected individuals at the same time. The results were in accordance with (Galtier & Gouy, 1998), where he and coworkers found diverse strains of this human entero virus in the water bodies of the city of Manaus in the Brazilian Amazon.

As far as test of the homogeneity of substitution patterns (Disparity Index Test) between sequences was concerned the assumed null hypothesis was "the analyzed TTV sequences have evolved by following the same nucleotide substitution pattern". The probability of rejecting or accepting the null hypothesis was calculated using disparity index test that involves Monte Carlo's simulation (500 replicates). *P*-value for significance level was set at 5%. For the isolated sequences of TTV in the present study the testing of this assumption was important before applying any molecular evolutionary analysis. Information of the violation of the homogeneity assumption would facilitate in choosing advanced methods of phylogenetic reconstruction (Galtier & Gouy, 1998; lockhart *et al.*, 1994) or the alternative option was the removal of uninformative sequences, if probable, before conducting phylogenetic analyses. Identification of viral genomes (ORF1 region in the present study) with uncommon patterns of evolutionary change is also advantageous for interpreting the evolutionary mechanisms accountable for the experiential variations. As arbitrated from the degree of variances in base composition, bias between sequences was there for four pairs of TTV isolates (Kumar & Gadagkar, 2001). The assessments of the disparity index per site were shown for each sequence pair. The analysis involved 17 nucleotide sequences, from which all

positions containing gaps and missing data were eliminated. There were a total of 270 positions in the final dataset. Four of the TTV viral pair sequences indicated low *P*-values smaller than 0.05, and were considered significant (Table 2, bold black). Hypothesis of homogeneity in nucleotide substitution was rejected in these four pair of sequences i.e., JX885496_Pak4 and JX885499_Pak7 (*p-value* 0.02), JX885499_Pak7 and JX885507_Pak15 (*p-value* 0.05), JX885502_Pak10 and JX885507_Pak15 (*p-value* 0.05), and JX885505_Pak13 and JX885507_Pak15 (*p-value* 0.006). Evolutionary analyses were conducted in MEGA5.10 Beta#4 (Tamura *et al.*, 2011). As four pairs of sequences were with heterogeneous substitution pattern, all remaining fulfilled the assumption; therefore they were not removed for the phylogenetic tree construction.

In normally constructed cladograms the evolutionary data is at the tips, nodes and branches, while the rest is imputation (Gould, 1977). Normally the species are stable, and they take millions of years to accumulate small changes, this flat phase is "punctuated" by a series of eruption of variations, resulting in a new form of species and thus it is named as punctuated equilibrium (Gould & Eldredge, 2000). Punctuated equilibrium connects the missing links of evolutionary processes (Caulin & Maley, 2011; McInerney 2013). Stressed that evolution of biological species is not gradual but there occurs an intermittent burst of activity and then a long period of quiescence for biological evolution to occur. Extinctions are undeniably, sporadic at all gauges and fitness landscape signifies the capacity of species to live as a function of their genetic code (Barnosky *et al.*, 2011; Raup, 1986).

Additionally, for evolution to take place there should be acceptance of only beneficial evolutionary shifts and these moves should be with low probability. The species are virtually always at confined fitness maxima (Bak & Sneppen, 1993) as can be experiential in phylogram with *p-distances* that an active evolution was observed in the species where positive immune pressure of the host was involved as shown in clade one of Figure 3. (JX885493_Pak1, JX885494_Pak2, JX885495_

Pak3, JX885496_Pak4, JX885499_Pak7, JX885500_Pak8, JX885501_Pak9, JX885502_Pak10, JX885503_Pak11, JX885504_Pak12, JX885505_Pak13, JX885507_Pak15 and JX885509_Pak17). Neutralizing antibodies are formed by the host according to viral strains, this in turn builds up a high immune pressure, the virus in turn mutates to ensure its existence (Farci *et al.*, 1994). Moreover there is a contribution of both host and viral factors in the stable or energetically mutating evolutionary pattern of the virus e.g., host's failure to raise an active immune response against the invading virus, heterogeneity of viral genetic makeup in terms of genogroups and genotypes and natural recombination of its genome (Irshad *et al.*, 2006; Irving *et al.*, 1999; Kenar Koohi *et al.*, 2012; Pinho-Nascimento *et al.*, 2011). As opposed to this a totally reverse trend of higher stability was observed for the species under less immune pressure clade II Figure 3 (JX885497_Pak5, JX885498_Pak6, JX885506_Pak14 and JX885508_Pak16). A comparative Phylogram generated by the sequences of TTV virus in current study and from retrieved sequences of nucleotide databases indicated that all the ORF1 sequences of the current study were the examples of simplicifolious clades (with only one leaf on the clade). Although they were related to each other but are evolving divergently (Figure 4). This was a general pattern observed for all other sequences except a few bifolious and trifolious clades from Thailand, Italy, Iran and China.

Tajima's test of neutrality uses nucleotide assortment to compare the number of discriminating nucleotides per site. A site is measured segregating if in a comparison of m number of sequences, at that site there are two or more nucleotides; the average number of nucleotide differences per site between two sequences will be the nucleotide diversity. If all the alleles are selectively neutral, then the product 4Nv (where N is actual population size and v is the mutation rate per site) can be assessed in two ways, and the difference in the estimate obtained provides an indication of non-neutral evolution (Tajima, 1989; Yang & Rannala, 2012). The positive D value of Tajima's test obtained in Table 1 indicated that the different strain polymorphism found in the ORF1 region of TTV was possibly upheld by a course of balancing selection, persistent by an evolutionary way of gene birth and gene death. Such combination has thus selected for the best polymorphism in accordance with TTV types by host selection and is an example of divergent evolution.

Concordant results were obtained by a study that was conducted on diverse members of kingdom Protista where the same evolutionary trend was observed on the basis of positive D value (Escobar & Castaño, 2009). Likewise if a negative D value is obtained it proposes very low polymorphism in selected genome, demonstrating a purifying selection process or an increase in viral population size, or a bottle neck effect in the beginning of a typical viral infection. On the contrary, positive values of Tajima's Neutrality test specify enormous polymorphism in selected TTV Isolates and compact viral population magnitude, thereby interceding a balancing selection process (Kimura & Ohta, 1971; Nei *et al.*, 2010).

As already mentioned that TTV isolates from Pakistan showed their independent existence (Branch and tree topologies), the isolates were different from each other and the ones reported from other parts of the world, but they formed a single cluster. The findings of the present study can be strengthened by a previous hypothesis, which states that in worldwide human population a whole swarm of TTV and TTV like viruses circulate. Regardless of this far-reaching sequence diversity, all variants of TTV have in common, a genomic organization (with three expected encoded proteins with functional and length similarities) (Irshad *et al.*, 2006).

More or less all of the TTV sequences in the present study, were capable of genetic adaptation according to immune responses of their hosts, hence their adaptation to heighten their fitness. This entire process is a sum of positive and negative selection in the form of functional control on immune recognition and replication. The results of the present study were in agreement with the ones reported from Iran (Keenar Koohi *et al.*, 2012). This extremely enhanced capability

of TTV adaptability according to changing immune responses is attributed to its circular genome and its DNA based genetic makeup, as RNA viruses mutate more rapidly than DNA viruses, surely there will be certain factors that are still unknown, in making TTV successful survivor globally in a range of cell and host tropism (Irshad *et al.*, 2006; Mankotia & Irshad, 2014; Peng *et al.*, 2015).

CONCLUSIONS

The results of the present study contradict the results from previous report from Pakistan. Low seroprevalence has been observed in the current study with high sequence variability in ORF1 region of TTV genome. These variations are more pronounced in open reading frames as compared to untranslated regions that are relatively conserved. Virus seems to be following a pattern of divergent evolution and has independently and successfully evolved by mutations and replications without the involvement of any other virus. There seems to be no association between HCV and TTV except the common modes of transmission being that they are both blood borne pathogens. Active mutations in the virus and rearrangements suggest high number of serological and genetic viral variants.

Compliance with ethical standards

Conflict of interest: The authors declare that they have no competing interests.

Ethical approval for research involving human participants: All Procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consents: Informed consent was obtained from all individual participants included in

REFERENCES

Abbas, D., Dehkordi, P.G. & Hajimirzaei, M.R. (2011). The prevalence of transfusion-transmitted virus (TTV) infection in patients with chronic hepatitis B and C in Southwest of Iran. *American Journal of Botany* **10**(25): 4954-4957.

Ali, S., Van Pelt, J.F., Verslype, C., Nevens, F., Fevery, J. & Yap, S.H. (2004). TT virus infection in acute and chronic liver diseases and in patients regularly receiving blood products in Belgium. *Acta Gastro-Enterologica Belgica* **67**(2): 161-165.

Armstrong, G.L., Wasley, A., Simard, E.P., McQuillan, G.M., Kuhnert, W.L. & Alter, M.J. (2006). The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. *Annals of Internal Medicine* **144**(10): 705–714. http://doi.org/10.7326/0003-4819-144-10-200605160-00004

Bak, P. & Sneppen, K. (1993). Punctuated equilibrium and criticality in a simple model of evolution. *Physical Review Letters* **71**(24): 4083-4086. http://doi.org/10.1103/PhysRevLett.71.4083

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., Mersey, B. & Ferrer, E.A. (2011). Has the Earth's sixth mass

extinction already arrived? *Nature* **471**(7336): 51-57. http://doi.org/10.1038/nature09678

Béland, K., Dore-Nguyen, M., Gagné, M.-J., Patey, N., Brassard, J., Alvarez, F. & Halac, U. (2014). Torque Teno virus in children who underwent orthotopic liver transplantation: new insights about a common pathogen. *The Journal of Infectious Diseases* **209**(2): 247-254. http://doi.org/10.1093/infdis/jit423

Bendinelli, M., Pistello, M., Maggi, F., Fornai, C., Freer, G. & Vatteroni, M.L. (2001). Molecular properties, biology, and clinical implications of TT virus, a recently identified widespread infectious agent of humans. *Clinical Microbiology Reviews* **14**(1): 98-113. http://doi.org/10.1128/CMR.14.1.98-113.2001

Bernardin, F., Operskalski, E., Busch, M. & Delwart, E. (2010). Transfusion transmission of highly prevalent commensal human viruses. *Transfusion* **50**(11): 2474-2483. http://doi.org/10.1111/j.1537-2995.2010.02699.x

Blazsek, A., Sillo, P., Ishii, N., Gergely, P., Poor, G., Preisz, K., Hashimoto, T., Medvecz, M. & Kárpáti, S. (2008). Searching for foreign antigens as possible triggering factors of autoimmunity: Torque Teno virus DNA prevalence is elevated in sera of patients with bullous pemphigoid. *Experimental Dermatology* **17**(5): 446-454. http://doi.org/10.1111/j.1600-0625.2007.00663.x

Caulin, A.F. & Maley, C.C. (2011). Peto's Paradox: evolution's prescription for cancer prevention. *Trends in Ecology & Evolution* **26**(4): 175-182. http://doi.org/10.1016/j.tree.2011.01.002

De Villiers, E.M., Kimmel, R., Leppik, L. & Gunst, K. (2009). Intragenomic rearrangement in TT viruses: A possible role in the pathogenesis of disease. *Current Topics in Microbiology and Immunology* **331**: 91-107. http://doi.org/10.1007/978-3-540-70972-5_6

Diniz-Mendes, L., De Paula, V.S., Luz, S.L.B. & Niel, C. (2008). High prevalence of human Torque teno virus in streams crossing the city of Manaus, Brazilian Amazon. *Journal of Applied Micro-*

biology **105**(1): 51-58. http://doi.org/10.1111/j.1365-2672.2007.03720.x

Erker, J.C., Leary, T.P., Desai, S.M., Chalmers, M.L. & Mushahwar, I.K. (1999). Analyses of TT virus full-length genomic sequences. *Journal of Generly Virology* **80**(Pt 7): 1743-1750. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10423143

Escobar, A.J.G. & Castaño, D.M. (2009). Evolutionary analysis for the functional divergence of the Spo0A protein: the key sporulation control element. *In Silico Biology* **9**(4): 149-162. http://doi.org/10.3233/isb-2009-0405

Evgeny, V.V., Trofimov, D.Y., Tonevitsky, A.G., Ilinsky, V.V., Korostin, D.O. & Rebrikov, D.V. (2009). Torque Teno Virus (TTV) distribution in healthy Russian population. *Virology Journal* **6**(134): 4. http://doi.org/doi:10.1186/1743-422X-6-134

Farci, P., Alter, H.J., Wong, D.C., Miller, R.H., Govindarajan, S., Engle, R., Shapiro, M. & Purcell, R.H. (1994). Prevention of hepatitis C virus infection in chimpanzees after antibody-mediated *in vitro* neutralization. *Proceedings of the National Academy of Sciences of the United States of America* **91**(16): 7792-7796. http://doi.org/10.1073/pnas.91.16.7792

Fuels, I. & Gut, G. (2012). Commensal Microbes: The Frenemy Within. *Cell* **151**(5): 923-925. http://doi.org/10.1016/j.cell.2012.11.010

Galtier, N. & Gouy, M. (1998). Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Molecular Biology and Evolution* **15**(7): 871-879. http://doi.org/10.1093/oxfordjournals.molbev.a025991

Görzer, I., Jaksch, P., Kundi, M., Seitz, T., Klepetko, W. & Puchhammer-Stöckl, E. (2015). Pre-Transplant Plasma Torque Teno Virus Load and Increase Dynamics after Lung Transplantation. *PLoS One* **10**(4): e0122975. https://doi.org/10.1371/journal.pone.0122975

Gould, S.J. (1977). *Ontogeny and Phylogeny. Ecology* (Vol. 28). http://doi.org/10.2307/2412825

Gould, S.J. & Eldredge, N. (2000). Punctuated equilibrium comes of age. *Nature* **366**: 223-227.

Hall, T., Biosciences, I. & Carlsbad, C. (2011). BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences* **2**(June): 60-61.

Hino, S. & Miyata, H. (2007). Torque teno virus (TTV): current status. *Reviews in Medical Virology* **17**(1): 45-57. http://doi.org/10.1002/rmv.524

Hussain, T., Manzoor, S., Waheed, Y., Tariq, H. & Hanif, K. (2012). Phylogenetic analysis of Torque Teno Virus genome from Pakistani isolate and incidence of co-infection among HBV/HCV infected patients. *Virology Journal* **9**(1): 1. http://doi.org/10.1186/1743-422X-9-320

Irshad, M., Joshi, Y.K., Sharma, Y. & Dhar, I. (2006). Transfusion transmitted virus: A review on its molecular characteristics and role in medicine. *World Journal of Gastroenterology* **12**(32): 5122-5134.

Irving, W.L., Ball, J.K., Berridge, S., Curran, R., Grabowska, A.M., Jameson, C.L.,Neal, K.R., Ryder, S.D. & Thomson, B.J. (1999). TT virus infection in patients with hepatitis C: Frequency, persistence, and sequence heterogeneity. *Journal of Infectious Diseases* **180**(1): 27-34. http://doi.org/10.1086/314825

Iver, L., Ancreas, P., Desai, M.M., Pal, R.B. & Banker, D.D. (2005). Molecular epidemiology and clinical implications of TT Virus (TTV) Infection in Indian subjects. *Journal of Clinical Gastro-enterology* **39**(5): 422-429. http://doi.org/10.1097/01.mcg.0000159219.93160.bc

Jin, P., Fang, Y., Zhao, X. & Peng. Y. (2015). New prevalence estimate of Torque Teno virus (TTV) infection in healthy population and patients with chronic viral hepatitis in Jiujiang, China. *Virologica Sinica* **30**(3): 218-220. Retrieved from http://link.springer.com/article/10.1007%2Fs12250-014-3531-x

Jocimara, C.M., Patrícia, K.S., Roger, H.Y., Maria, A.E.W., Waldir, V. da. S.J., Alessandra, C.G.M. & Sueli, D.B. (2015). Prevalence of Torque teno virus in healthy donors of Paraná State, southern Brazil. *Revista Brasileira de Hematologia E Hemoterapia* **37**(5): 136-140. http://doi.org/10.1016/j.bjhh.2015.07.005

Jones, M.S., Kapoor, A., Lukashov, V.V, Simmonds, P., Hecht, F. & Delwart, E. (2005). New DNA viruses identified in patients with acute viral infection syndrome. *Journal of Virology* **79**(13): 8230-8236. http://doi.org/10.1128/JVI.79.13.8230-8236.2005

Karen, B. de. O. (2015). Torque teno virus: a ubiquitous virus. *Revista Brasileira de Hematologia E Hemoterapia* **37**(6): 357-358. http://doi.org/doi:10.1016/j.bjhh.2015.07.009

Kenar Koohi, A., Ravanshad, M., Rasouli, M., Falahi, S. & Baghban, A. (2012). Phylogenetic analysis of torque teno virus in hepatitis C virus infected patients in shiraz. *Hepatitis Monthly* **12**(7): 437-441. http://doi.org/10.5812/hepatmon.6133

Kheradpezhouh, M., Taremi, M., Gachkar, L., Aghabozorgi, S. & Khoshbaten, M. (2007). Presence and significance of transfusion-transmitted virus infection in Iranian patients on maintenance hemodialysis. *Journal of Microbiology, Immunology, and Infection* **40**(2): 106-111.

Kim, S.R., Hayashi, Y., Kudo, M., Imoto, S., Song, K.B., Ando, K., Shintani, S., Koterazawa, T., Kim, K.I. & Taniguchi, M. (2000). TTV positivity and transfusion history in non-B, non-C hepatocellular carcinoma compared with HBV- and HCV-positive cases. *Intervirology* **43**(1): 13–15. http://doi.org/int43013 [pii]

Kimura, M. & Ohta, T. (1971). Protein polymorphism as a phase of molecular evolution. *Nature* **229**(5285): 467-469. http://doi.org/10.1038/229467a0

Kumar, S. & Gadagkar, S.R. (2001). Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. *Genetics* **158**(3): 1321-1327.

Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T. J. &

Higgins, D. (2007). ClustalW and ClustalX version 2. *Bioinformatics* **23**(21): 2947-2948. http://doi.org/doi:10.1093/bioinformatics/btm404

Lockhart, P.J., Steel, M.A., Hendy, M.D. & Penny, D. (1994). Recovering evolutionary trees under a more realistic model of sequence evolution. *Molecular Biology and Evolution* **11**(4): 605-612. http://doi.org/{desktop}

Maggi, F., Pifferi, M., Fornai, C., Andreoli, E., Tempestini, E., Vatteroni, M., Presciuttini, S., Marchi, S., Pietrobelli, A., Boner, A., Pistello, M. & Bendinelli, M. (2003a). TT Virus in the Nasal Secretions of Children with Acute Respiratory Diseases/: Relations to Viremia and Disease Severity. *Journal of Virology* **77**(4): 2418-2425. http://doi.org/10.1128/JVI.77.4.2418

Maggi, F., Pifferi, M., Tempestini, E., Fornai, C., Lanini, L., Andreoli, E., Vatteroni, M., Presciuttini, S., Pietrobelli, A., Boner, A., Pistello, M. & Bendinelli, M. (2003b). TT virus loads and lymphocyte subpopulations in children with acute respiratory diseases. *Journal of Virology* **77**(16): 9081-9083.

Mankotia, D.S. & Irshad, M. (2014). Cloning and expression of N22 region of Torque Teno virus (TTV) genome and use of peptide in developing immunoassay for TTV antibodies. *Virology Journal* **11**(1): 96. http://doi.org/10.1186/1743-422X-11-96

Mauss, S., Berg, T., Rockstroh, J., Sarrazin, C. & Wedemeyer, H. (2017). *HEPATOLOGY. A clinical textbook.* 8th Edition. Published by Medizin Fokus Verlag. Germany.

Maximova, N., Pizzol, A., Ferrara, G., Maestro, A. & Tamaro, P. (2015). Does teno torque virus induce autoimmunity after hematopoietic stem cell transplantation? A case report. *Journal of Pediatric Hematology/Oncology* **37**(3): e194-e197. http://doi.org/10.1097/MPH.0000000000000194

McInerney, J.O. (2013). More than tree dimensions: Inter-lineage evolution's ecological importance. *Trends in Ecology and Evolution* **28**: 924-625. http://doi.org/10.1016/j.tree.2013.09.002

Nei, M., Suzuki, Y. & Nozawa, M. (2010). The neutral theory of molecular evolution in the genomic era. *Annual Review of Genomics and Human Genetics* **11**(June): 265-289. http://doi.org/10.1146/annurev-genom-082908-150129

Neumann, R.S., Kumar, S. & Shalchian-Tabrizi, K. (2014). BLAST output visualization in the new sequencing era. *Briefings in Bioinformatics* **15**(4): 484-503. http://doi.org/10.1093/bib/bbt009

Nishizawa, T., Okamoto, H., Konishi, K., Yoshizawa, H., Miyakawa, Y. & Mayumi, M. (1997). A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochemical and Biophysical Research Communications* **241**(1): 92-97. http://doi.org/10.1006/bbrc.1997.7765

Okamoto, H., Takahashi, M., Nishizawa, T., Ukita, M., Fukuda, M., Tsuda, F., Miyakawa, Y. & Mayumi, M. (1999). Marked genomic heterogeneity and frequent mixed infection of TT virus demonstrated by PCR with primers from coding and noncoding regions. *Virology* **259**(2): 428-436. http://doi.org/10.1006/viro.1999.9770

Peng, Y.H., Nishizawa, T., Takahashi, M., Ishikawa, T., Yoshikawa, A. & Okamoto, H. (2002). Analysis of the entire genomes of thirteen TT virus variants classifiable into the fourth and fifth genetic groups, isolated from viremic infants. *Archives of Virology* **147**(1): 21-41.

Pinho-Nascimento, C.A., Leite, J.P.G., Niel, C. & Diniz-Mendes, L. (2011). Torque Teno Virus in Fecal Samples of Patients With Gastroenteritis: Prevalence, Genogroups Distribution, and Viral Load. *Journal of Medical Virology* **83**(6): 1107-1111. http://doi.org/10.1002/jmv.22024

Plamondon, M., Labbe, A.C., Frost, E., Deslandes, S., Alves, A.C., Bastien, N. & Pepin, J. (2007). Hepatitis C virus infection in Guinea-Bissau: a sexually transmitted genotype 2 with parenteral amplification? *PLoS One* **2**(4): e372. http://doi.org/10.1371/journal.pone.0000372

Prescott, L.E., MacDonald, D.M., Davidson, F., Mokili, J., Pritchard, D.I., Arnot, D.E., Riley, E.M., Greenwood, B.M., Hamid, S., Saeed, A.A., McClure, M.O., Smith, D.B. & Simmonds, P. (1999). Sequence diversity of TT virus in geographically dispersed human populations. *Journal of Genral Virology* **80**(Pt 7): 1751-1758.

Raup, D. (1986). Biological extinction in earth history. *Science* **231**(4745): 1528-1533. http://doi.org/10.1126/science.11542058

Simmonds, P., Davidson, F., Lycett, C., Prescott, L.E., MacDonald, D.M., Ellender, J., Yap, P.L., Ludlam, C.A., Haydon, G.H., Gillon, J. & Jarvis, L.M. (1998). Detection of a novel DNA virus (TTV) in blood donors and blood products. *Lancet* **352**(9123): 191-195. http://doi.org/S0140673698030566 [pii]

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**(3): 585-595. http://doi.org/PMC1203831

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**(10): 2731-2739. http://doi.org/10.1093/molbev/msr121

Thomas, R.H. (2001). Molecular Evolution and Phylogenetics. *Heredity* (Vol. 86). http://doi.org/10.1046/j.1365-2540.2001.0923a.x

Tokita, H., Murai, S., Kamitsukasa, H., Yagura, M., Harada, H., Takahashi, M. & Okamoto, H. (2002). High TT virus load as an independent factor associated with the occurrence of hepatocellular carcinoma among patients with hepatitis C virus-related chronic liver disease. *Journal of Medical Virology* **67**(4): 501-509.

Yang, Y. & Yooseph, S. (2013). SPA: A short peptide assembler for metagenomic data. *Nucleic Acids Research* **41**(8): 1-10. http://doi.org/10.1093/nar/gkt118

Yang, Z. & Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nature Reviews Genetics* **13**(5): 303-314. http://doi.org/10.1038/nrg3186