RESEARCH   **Open Access**

Decoding the codon usage patterns in Y-domain region of hepatitis E viruses

Zoya Shafat[1], Anwar Ahmed[2], Mohammad K. Parvez[3] and Shama Parveen[1*]

Abstract

Background: Hepatitis E virus (HEV) is a   positive-sense RNA virus belonging to the family *Hepeviridae*. The genome of HEV is organized into three open-reading frames (ORFs): ORF1, ORF2, and ORF3. The ORF1 non-structural Y-domain region (YDR) has been demonstrated to play an important role in the HEV pathogenesis. The nucleotide composition, synonymous codon usage bias in conjunction with other factors influencing the viral YDR genes of HEV have not been studied. Codon usage represents a significant mechanism in establishing the host-pathogen relationship. The present study for the first time elucidates the detailed codon usage patterns of YDR among HEV and HEV-hosts (Human, Rabbit, Mongoose, Pig, Wild boar, Camel, Monkey).

Results: The overall nucleotide composition revealed the abundance of C and U nucleotides in YDR genomes. The relative synonymous codon usage (RSCU) analysis indicated biasness towards C and U over A and G ended codons in HEV across all hosts. Codon frequency comparative analyses among HEV-hosts showed both similarities and discrepancies in usage of preferred codons encoding amino acids, which revealed that HEV codon preference neither completely differed nor completely showed similarity with its hosts. Thus, our results clearly indicated that the synonymous codon usage of HEV is a mixture of the two types of codon usage: coincidence and antagonism. Mutation pressure from virus and natural selection from host seems to be accountable for shaping the codon usage patterns in YDR. The study emphasised that the influence of compositional constraints, codon usage biasness, mutational alongside the selective forces were reflected in the occurrence of YDR codon usage patterns.

Conclusions: Our study is the first in its kind to have reported the analysis of codon usage patterns on a total of seven different natural HEV hosts. Therefore, knowledge of preferred codons obtained from our study will not only augment our understanding towards molecular evolution but is also envisaged to provide insight into the efficient viral expression, viral adaptation, and host effects on the HEV YDR codon usage.

Keywords: YDR, Nucleotide composition, Codon usage bias, Mutation pressure, Natural selection

Background

Hepatitis E virus (HEV) is the cause of both epidemic and sporadic hepatitis cases in humans [1, 2]. HEV is a positive-sense, single-stranded RNA virus, belonging to the family *Hepeviridae*. The 7.2 kb genome of HEV, with short 5′ and 3′ non-coding regions (NCR), consists of three partially overlapping open reading frames (ORFs) [3]. The 5′ most ORF (ORF1) encodes the non-structural polyprotein which is organized into seven functional domains including the Y-domain region (YDR) [4, 5], 3′ most ORF (ORF2) codes for the viral capsid protein [6, 7], and ORF3 encodes the phosphoprotein responsible for viral regulation [8–10]. The non-structural ORF1 Y-domain region (YDR) critical residues have been demonstrated to play critical role in the HEV life cycle [11].

HEV is segregated into four major genotypes (HEV-1 to HEV-4), out of which HEV-1 and HEV-2 infect humans, while HEV-3 and HEV-4 strains have an expanded range of hosts which includes humans, rabbits, wild boars, and

*Correspondence: sparveen2@jmi.ac.in; shamp25@yahoo.com
[1] Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India
Full list of author information is available at the end of the article

pigs [12–25]. Studies have reported the isolation of other strains of HEV from specific hosts, such as HEV-5 and HEV-6 from wild boars in Japan [14, 15]. HEV-7 from dromedary camels [24] and HEV-8 from Bactrian camels [25]. The genetic code encompasses 64 codons, separated into 20 distinguishable groups. Each individual group, consists of one to six codons, encodes the same amino acid. Thus, each standard amino acid is often encoded by alternative codons belonging to the same group. These alternative codons are termed as "synonymous" codons. These synonymous codons differ not only between genomes but also within the same genome of the organisms/organism. This phenomenon is referred to as codon usage bias [26, 27] and has been well documented in many organisms including prokaryotes, eukaryotes, and viruses [28–33].

Previous reports on codon usage have determined various factors governing the codon usage patterns which include mutational pressure, translational selection, G + C content secondary structure of protein, selective transcription replication, hydrophilicity, and hydrophobicity of the protein and the external environment [28, 34–36]. Among these, compositional constraints under natural selection and mutational pressure are two major paradigms in shaping the codon usage patterns in organisms [37–39]. However, in viruses, mutational pressure rather than natural selection is found to be the major factor influencing codon usage variation [40–43].

As YDR indispensability in HEV pathogenesis has been demonstrated [11], thus, it is important to determine the distinctive genetic features that are prevalent in their genomes. Using an interdisciplinary systems biology approach, we attempted to explain the codon usage bias of HEV-hosts in conjunction with evolutionary forces (compositional, mutational, selection) accountable for shaping the YDR codon usage patterns. The present study is the first in its kind which have reported the detailed codon usage analysis on a total of 7 hosts in HEV YDR. Therefore, knowledge obtained from the presented study will not only augment our understanding towards molecular evolution but is also envisaged to provide insight into the efficient viral expression, viral adaptation and host effects on HEV [31, 44].

Methods
Heat map construction
The heat map was constructed using the online software tool Morpheus (https://software.broadinstitute.org/morpheus/documentation.html). Heat map is one of the most commonly used visualization in the science field because it allows us to find patterns in our data, compact a large amount of information into a small space, and are a natural representation of a matrix.

Sequence data acquisition
The YDR sequences were accumulated from the National Centre for Biotechnology information (NCBI). The retrieved sequences were selected based on the following inclusion criteria: (a) The strain (GenBank Accession number: NC_001434.1) was used as reference strain; (b) sequences were included from different hosts encompassing human, rabbit, pig, mongoose, wild boar, camel, and monkey; (c) sequences from same or different regions at varying time intervals were considered to avoid repetition in analysis; and (d) sampling dates of the sequences were clearly stated. Accumulated sequences from NCBI were edited using the Bioedit v.7.2 sequence analysis software (http://bioedit.software.informer.com/7.2/). The sequences were further manually edited to exclude ambiguous portions to obtained non-structural ORF1 gene product YDR before proceeding for the final alignment. Multiple alignments for YDR sequences datasets were carried out using Clustal X2 Algorithm (http://www.clustal.org/clustal2/) [17]. The complete list of the sequences used for various host organisms are listed as additional files in the supplementary information (Additional file 1: S1 Table, Additional file 2: S2 Table, Additional file 3: S3 Table, Additional file 4: S4 Table, Additional file 5: S5 Table, Additional file 6: S6 Table, Additional file 7: S7 Table, Additional file 8: S8 Table).

Nucleotide composition analysis
Nucleotide composition analysis of the YDR was calculated using MegaX software. The overall nucleotides occurrence frequency (A%, C%, T/U%, and G%), overall occurrence of nucleotide frequency at the third position of codon (A3%, C3%, U3%, and G3%) and overall occurrence of nucleotides frequencies of G+C at different codon positions were determined. The AUG and UGG codons were not considered for the analysis as they do not exhibit codon usage bias. The termination codons (UAG, UGA, UAA) were also excluded from the analysis since they do not encode any amino acid.

Relative synonymous codon usage (RSCU) analysis
The ratio between the observed and expected usage frequency of a codon is described as the RSCU value if all synonymous codons are used equally for any specific amino acid [18]. The RSCU index was determined as follows:

$$RSCU = \frac{Gij}{\sum_{j}^{ni} Gij} ni$$

where *RSCU* is the relative synonymous codon usage value, G_{ij} is the observed number of the *i*th codon for the *j*th amino acid that has an "ni" type of synonymous

Shafat *et al. Journal of Genetic Engineering and Biotechnology* (2022) 20:56

Page 3 of 14

codon. The RSCU values of the YDR were calculated using MegaX to determine the codon usage characteristics without the effect of amino acid composition and coding sequence length. Codons with RSCU values (> 1.6) and (< 0.6) were considered as "over-represented" and "under-represented" codons, respectively, whereas codons having the RSCU values (1) were regarded as not biased (average level codon). Moreover, less-abundant (RSCU < 1) and more-abundant (RSCU > 1) used codons were also determined.

Relationship between overall nucleotide composition and nucleotide composition at the 3rd codon position

The correlation between A, T, G, C, GC, and 3rd codon position of its counterparts (A3, T3, G3, C3, GC3) were assessed. This was carried out to analyze whether if natural selection/mutation pressure individually contributed or if both collaboratively influenced the evolution of YDR in HEVs.

Results

Compositional features of YDR

The nucleotide composition values for YDR were calculated to analyze the effect of compositional constraints on codon usage (Table 1) (Fig. 1).

HEV

The nucleotide composition trend was in order C > U > G > A, with an average of 30.169%, 26.631%, 24.357%, and 18.841%, respectively. Synonymous codons at the third

position followed the trend C3S > U3S > G3S > A3S. The overall GC content was higher than that of AU, with 54.526% observed, compared with 45.472%, respectively, which indicates a GC-biased composition (Additional file 1: S1 Table).

Human

The nucleotide composition trend was in order C > U > G > A, with an average of 28.022%, 27.654%, 25.003%, and 19.319%, respectively. Synonymous codons at the third position followed the trend U3S > C3S > G3S > A3S. The overall GC content was higher than that of AU, with 53.025% observed, compared with 46.973%, respectively, which indicates a GC-biased composition (Additional file 2: S2 Table).

Rabbit

The nucleotide composition trend was in order C > U > G > A, with an average of 29.816%, 27.777%, 24.277%, and 18.127%, respectively. Synonymous codons at the third position followed the trend C3S > U3S > G3S > A3S. The overall GC content was higher than that of AU, with 54.093% observed, compared with 45.904%, respectively, which indicates a GC-biased composition (Additional file 3: S3 Table).
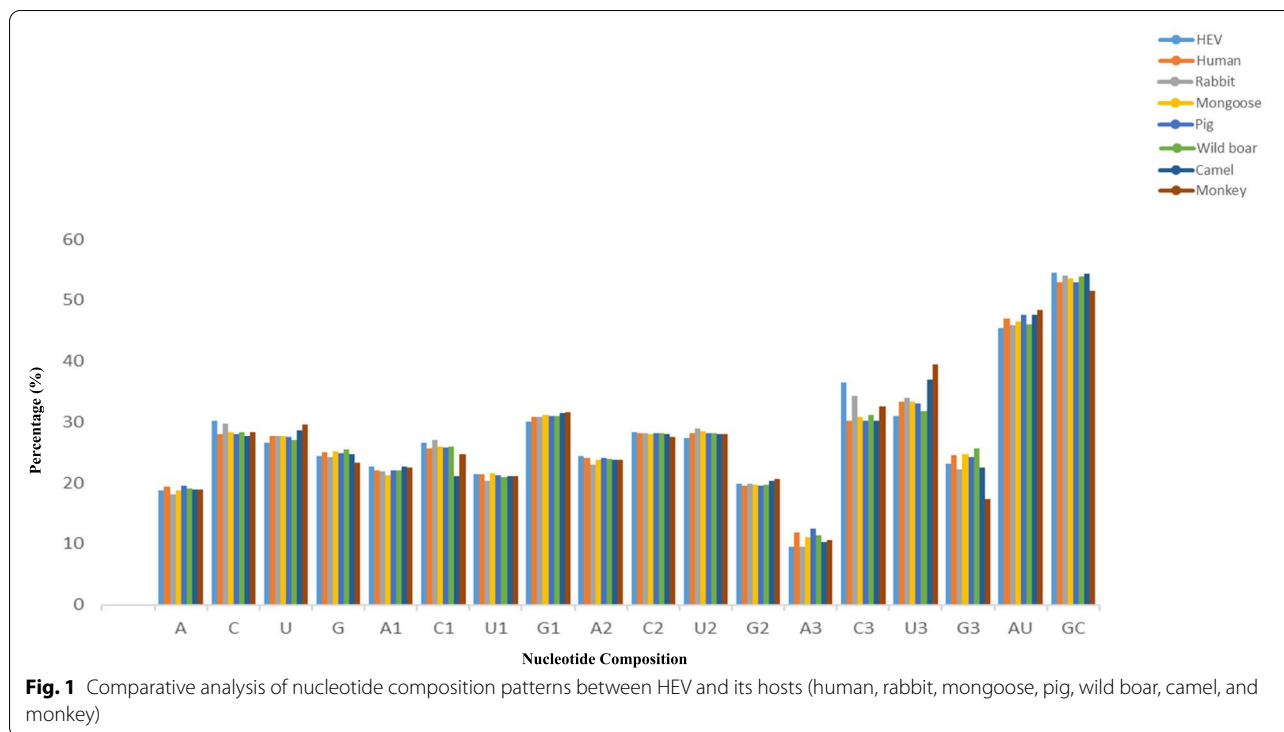
Mongoose

The nucleotide composition trend was in order C > U > G > A, with an average of 28.287%, 27.777%, 25.229%, and 18.705%, respectively. Synonymous codons at the third

Table 1 Nucleotide composition analysis of YDR of hepatitis E viruses (%)

Nucleotide	HEV	Human	Rabbit	Mongoose	Pig	Wild boar	Camel	Monkey
A	18.841	19.319	18.127	18.705	19.532	19.108	18.910	18.960
C	30.169	28.022	29.816	28.287	28.048	28.391	27.662	28.287
U	26.631	27.654	27.777	27.777	27.485	27.014	28.671	29.510
G	24.357	25.003	24.277	25.229	24.933	25.485	24.755	23.241
A1	22.698	22.057	21.844	21.253	21.997	21.990	22.660	22.477
C1	26.665	25.688	27.013	25.993	25.792	25.908	21.100	24.770
U1	21.505	21.403	20.387	21.559	21.246	21.037	21.100	21.100
G1	30.131	30.851	30.784	31.192	30.963	31.062	31.467	31.651
A2	24.349	24.077	23.037	23.853	24.045	23.888	23.713	23.853
C2	28.387	28.186	28.134	27.981	28.175	28.158	27.941	27.522
U2	27.408	28.147	29.001	28.440	28.216	28.228	28.033	27.981
G2	19.853	19.588	19.826	19.724	19.562	19.724	20.312	20.642
A3	9.480	11.821	9.531	11.009	12.554	11.440	10.366	10.550
C3	36.452	30.193	34.301	30.886	30.177	31.108	30.275	32.568
U3	30.978	33.414	33.944	33.333	32.992	31.779	36.880	39.449
G3	23.088	24.570	22.222	24.770	24.275	25.670	22.477	17.431
AU	45.472	46.973	45.904	46.482	47.617	46.122	47.581	48.47
GC	54.526	53.025	54.093	53.516	52.981	53.876	54.417	51.528
GC3	59.54	54.763	56.523	55.656	54.452	56.778	52.752	49.999

The values are represented as percentage

Shafat *et al. Journal of Genetic Engineering and Biotechnology* (2022) 20:56

Page 4 of 14

Fig. 1 Comparative analysis of nucleotide composition patterns between HEV and its hosts (human, rabbit, mongoose, pig, wild boar, camel, and monkey)

position followed the trend U3S > C3S > G3S > A3S. The overall GC content was higher than that of AU, with 53.516% observed, compared with 46.482%, respectively, which indicates a GC-biased composition (Additional file 4: S4 Table).

Pig
The nucleotide composition trend was in order C > U > G > A, with an average of 28.048%, 27.485%, 24.933%, and 19.532%, respectively. Synonymous codons at the third position followed the trend U3S > C3S > G3S > A3S. The overall GC content was higher than that of AU, with 52.981% observed, compared with 47.617% respectively, which indicates a GC-biased composition (Additional file 5: S5 Table).

Wild boar
The nucleotide composition trend in HEV was in order C > U > G > A, with an average of 28.391%, 27.014%, 25.485%, and 19.108%, respectively. Synonymous codons at the third position followed the trend U3S > C3S > G3S > A3S. The overall GC content was higher than that of AU, with 53.876% observed, compared with 46.122%, respectively, which indicates a GC-biased composition (Additional file 6: S6 Table).

Camel
The nucleotide composition trend in HEV was in order U > C > G > A, with an average of 28.671%, 27.662%,

24.755%, and 18.910%, respectively. Synonymous codons at the third position followed the trend U3S > C3S > G3S > A3S. The overall GC content was higher than that of AU, with 54.417% observed, compared with 47.581 respectively, which indicates a GC-biased composition (Additional file 7: S7 Table).

Monkey
The nucleotide composition trend in HEV was in order U > C > G > A, with an average of 29.510%, 28.287%, 23.241%, and 18.960%, respectively. Synonymous codons at the third position followed the trend U3S > C3S > G3S > A3S. The overall GC content was higher than that of AU, with 51.528% observed, compared with 48.47%, respectively, which indicates a GC-biased composition (Additional file 8: S8 Table).

Thus, the overall initial compositional findings revealed that YDR was richly endowed with C and U nucleotides. It was observed that the least chosen nucleotide in YDR was A. Moreover, the GC contents were significantly higher than that of AU contents (since AT content was <50%) in YDR.

Patterns of codon usage in YDR
RSCU analysis was performed to assess the codon usage patterns and preferences for synonymous codons in the YDR. The RSCU values were computed for every codon in each gene sequence to decrypt the extent to which

C/U-ended codons were preferred. The results are mentioned in Table 2 (Fig. 2).

HEV
Among the 29 preferred codons, 24 were U/C-ending (U-ending: 12; C-ending: 12;) and 5 were G/A-ending (G-ending: 5; A-ending: 0) (Table 2) (Additional file 9: S9 Table). This result inferred that U- and C-ending codons are preferred in coding sequences. Within these preferred codons, 13 had a RSCU value >1.6, i.e., overrepresented codons (CUU, CUC, UCC, CCU, ACC, GCC, CAG, AAG, GAG, UGC, CGU, CGC, GGC), while the remaining 16 had RSCU values >0.6 and <1.6. Out of these 16, 4 codons had RSCU values <1, i.e., less-abundant codons (UUC, GUG UAU, GAU), and 12 had RSCU values >1, i.e., abundant codons (UUU, AUU, AUC, GUU, GUC, ACU, GCU, UAC, CAU, GAC, CGG, GGU). No optional synonymous codons were underrepresented (RSCU < 0.6).

Human
Among the 28 preferred codons, 23 were U/C-ending (U-ending: 13; C-ending: 10;) and 5 were G/A-ending (G-ending: 5; A-ending: 0) (Table 2) (Additional file 10: S10 Table). This result inferred that U- and C-ending codons are preferred in coding sequences. Within these preferred codons, 6 had a RSCU value >1.6, i.e., overrepresented codons (CUU, CUC, UCU, GAG, CGU, GGC), while the remaining 22 had RSCU values >0.6 and <1.6. Out of these 22, 4 codons had RSCU values <1, i.e., less-abundant codons (GCU, GCG, UAU, GAC), while 18 had RSCU values >1, i.e., abundant codons (UUU, CUG, AUU, AUC, GUU, GUC, GUG, UCC, CCU, ACU, ACC, GCC, UAC, CAU, AAG, GAU, CGC, GGU). No optional synonymous codons were underrepresented (RSCU < 0.6).

Rabbit
Among the 29 preferred codons, 23 were U/C-ending (U-ending: 12; C-ending: 11) and 6 were G/A-ending (G-ending: 5; A-ending: 1) (Table 2) (Additional file 11: S11 Table). This result inferred that U- and C-ending codons are preferred in coding sequences. Within these preferred codons, 7 had RSCU value >1.6, i.e., overrepresented codons (CUU, CUC, GUC, UCU, CAG, CGU, CGC), while the remaining 22 preferred codons had RSCU values >0.6 and <1.6. Out of these 22, 3 codons had RSCU values <1, i.e., less-abundant codons (UUU, GCG, UAC), while 19 had RSCU values >1, i.e., abundant codons (UUC, CUG, AUU, AUC, GUU, GUG, CCC, ACU, ACC, GCU, GCC, UAU, CAU, AAA, GAU, GAG, UGC, GGU, GGC). No optional synonymous codons were underrepresented (RSCU < 0.6).

Mongoose
Among the 29 preferred codons, 21 were U/C-ending (U-ending: 12; C-ending: 9) and 8 were G/A -ending (G-ending: 7; A-ending: 1) (Table 2) (Additional file 12: S12 Table). This result inferred that U-and C-ending codons are preferred in coding sequences. Within these preferred codons, 10 had RSCU value >1.6, i.e., overrepresented codons (CUC, GUC, UCU, CCU, ACU, CAG, GAG, CGU, CGC, GGC) while the remaining 19 preferred codons had RSCU values >0.6 and <1.6. Out of these 19, 3 codons had RSCU values <1, i.e., less-abundant codons (GUG, GCG, UAC), while 16 had RSCU values >1, i.e., abundant codons (UUU, CUU, CUG, AUC, AUA, GUU, ACC, GCU, GCC, UAU, CAU, AAG, GAU, UGC, CGG, GGU). One optional synonymous codon was underrepresented (RSCU < 0.6).

Pig
Among the 25 preferred codons, 20 preferred codons were U/C-ending (U-ending: 13; C-ending: 7) and 6 were G/A -ending (G-ending: 4; A-ending: (1) (Table 1) (Additional file 13: S13 Table). This result inferred that U-and C-ending are preferred in coding sequences. Within these preferred codons, 8 had RSCU value >1.6, i.e., overrepresented codons (CUU, CUC, UCU, GCC, CAG, GAG, CGU, GGC), while the remaining 17 preferred codons had RSCU values >0.6 and <1.6. Out of these 17, 4 codons had RSCU values <1, i.e., less-abundant codons (AUC, GCU, GCA, UAC), while 13 had RSCU values >1, i.e., abundant codons (UUU, AUU, GUU, GUC, GUG, CCU, ACU, ACC, UAU, CAU, AAG, GAU, GGU). No optional synonymous codon was underrepresented (RSCU < 0.6).

Wild boar
Among the 27 preferred codons, 21 preferred codons were C/U-ending (C-ending: 11; U-ending: 10) and 6 were G/A -ending (G-ending: 5; A-ending: 1) (Table 1) (Additional file 14: S14 Table). This result inferred that C- and U-ending are preferred in YDR. Within these preferred codons, 8 had RSCU value >1.6, i.e., overrepresented codons (CUU, CUC, UCU, GAG, UGC, CGU, CGC, GGC), while the remaining 19 preferred codons had RSCU values >0.6 and <1.6. Out of these 19, 3 codons had RSCU values <1, i.e., less-abundant codons (AUC, AUA, UAC), while 16 had RSCU values >1, i.e., abundant codons (UUU, CUG, AUU, GUU, GUC, GUG, UCC, ACU, ACC, GCU, GCC, UAU, CAU, CAG, AAG, GAC). No optional synonymous codon was underrepresented (RSCU < 0.6).

Camel
Among the 24 preferred codons, 20 preferred codons were U/C-ending (U-ending: 13; C-ending: 7) and 4 were G/A -ending (G-ending: 3; A-ending: 1) (Table 2) (Additional file 15: S15 Table). This result inferred that U- and C-ending are preferred in YDR. Within these preferred codons, 8 had RSCU value >1.6, i.e., overrepresented
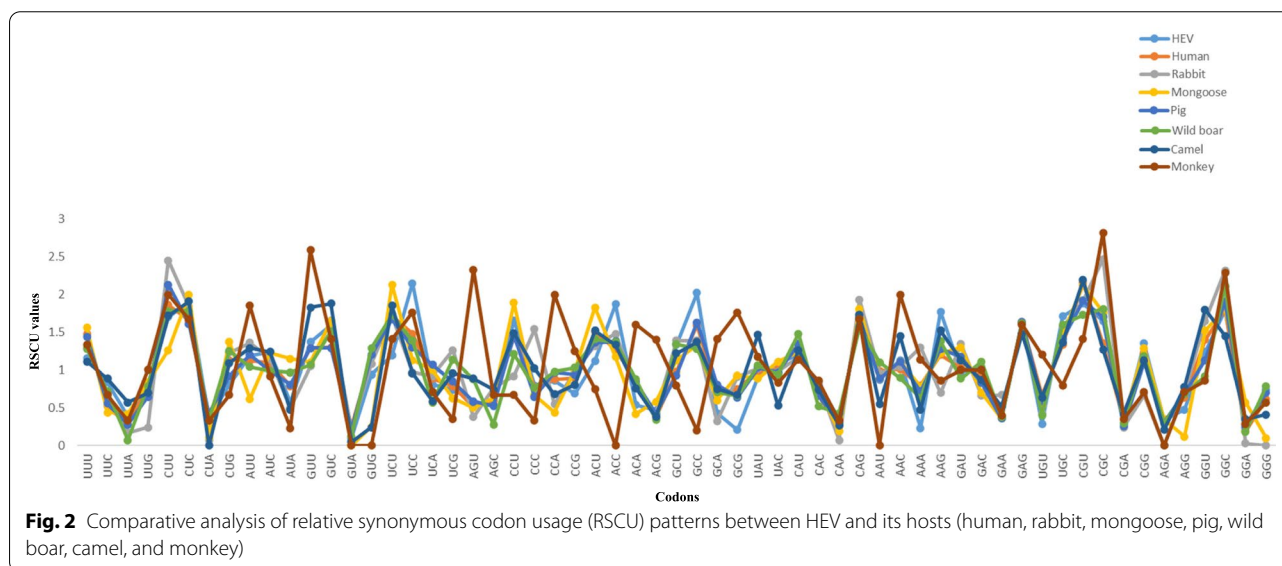
Shafat *et al. Journal of Genetic Engineering and Biotechnology*   (2022) 20:56

Page 6 of 14

Table 2 Average RSCU values of the codons of the HEV YDR and comparison with the RSCU values of its natural hosts

AA	Codon	HEV	Human	Rabbit	Mongoose	Pig	Wild boar	Camel	Monkey
Phe (F)	UUU	**1.16**	**1.48**	**1.3**	**1.56**	**1.44**	**1.28**	**1.11**	1.33
	UUC	**0.84**	0.52	**0.7**	0.44	0.56	0.72	**0.89**	0.67
Leu (L)	UUA	0.39	0.28	0.17	0.42	0.28	0.07	0.57	0.33
	UUG	0.81	0.77	0.24	0.84	0.65	0.79	0.7	1
	CUU	**1.7**	**1.86**	**2.45**	**1.26**	**2.13**	**1.78**	**1.72**	**2**
	CUC	**1.91**	**1.66**	**1.84**	**2**	**1.61**	**1.78**	**1.91**	**1.67**
	CUA	0.37	0.29	0.07	0.11	0.42	0.34	0	0.33
	CUG	0.81	**1.13**	**1.23**	**1.37**	0.92	**1.25**	1.09	0.67
Ile (I)	AUU	**1.19**	**1.12**	**1.36**	0.62	**1.2**	**1.04**	**1.29**	**1.85**
	AUC	**1.24**	**1.09**	**1.15**	**1.23**	**0.99**	**0.99**	**1.24**	**0.92**
	AUA	0.57	0.79	0.5	**1.15**	0.81	**0.97**	0.48	0.23
Val (V)	GUU	**1.37**	**1.29**	**1.05**	**1.1**	**1.29**	**1.07**	**1.83**	**2.59**
	GUC	**1.59**	**1.3**	**1.66**	**1.65**	**1.29**	**1.51**	**1.88**	**1.41**
	GUA	0.1	0.25	0.22	0	0.22	0.13	0.05	0
	GUG	**0.94**	**1.17**	**1.08**	0.25	**1.2**	**1.29**	0.24	0
Ser (S)	UCU	1.19	**1.72**	**1.72**	**2.13**	**1.67**	**1.73**	**1.85**	**1.41**
	UCC	**2.15**	**1.48**	0.97	1.13	1.3	**1.39**	0.96	**1.76**
	UCA	0.81	0.89	0.92	1	1.07	0.57	0.59	0.71
	UCG	0.74	0.79	1.26	0.63	0.85	1.14	0.96	0.35
	AGU	0.59	0.51	0.38	0.5	0.58	0.88	0.89	**2.33**
	AGC	0.52	0.6	0.76	0.63	0.53	0.28	0.74	0.67
Pro (P)	CCU	**1.66**	**1.44**	0.92	**1.89**	**1.45**	1.21	**1.49**	0.67
	CCC	0.8	0.78	**1.54**	0.67	0.65	0.77	1.02	0.33
	CCA	0.86	0.88	0.55	0.44	0.97	0.98	0.68	**2**
	CCG	0.69	0.9	0.99	1	0.94	1.03	0.81	**1.25**
Thr (T)	ACU	**1.12**	**1.5**	**1.32**	**1.83**	**1.36**	**1.41**	**1.52**	0.75
	ACC	**1.87**	**1.27**	**1.48**	**1.17**	**1.38**	**1.37**	**1.33**	0
	ACA	0.54	0.85	0.77	0.42	0.84	0.88	0.76	**1.6**
	ACG	0.47	0.38	0.42	0.58	0.42	0.34	0.38	**1.4**
Ala (A)	GCU	**1.33**	0.98	**1.39**	**1.13**	0.93	**1.34**	**1.22**	0.8
	GCC	**2.02**	**1.59**	**1.39**	**1.33**	**1.63**	**1.28**	**1.37**	0.2
	GCA	0.43	0.68	0.32	0.6	**0.81**	0.71	**0.75**	1.41
	GCG	0.21	**0.75**	**0.91**	**0.93**	0.64	0.66	0.67	**1.76**
Tyr (Y)	UAU	**0.91**	**0.99**	**1.02**	**0.89**	**1.03**	**1.06**	**1.47**	**1.17**
	UAC	**1.09**	**1.01**	**0.98**	**1.11**	**0.97**	**0.94**	0.53	**0.83**
His (H)	CAU	**1.2**	**1.34**	**1.17**	**1.24**	**1.34**	**1.48**	**1.26**	**1.14**
	CAC	0.8	0.66	0.83	0.76	0.66	0.52	0.74	0.86
Gln (Q)	CAA	0.36	0.27	0.07	0.19	0.31	0.43	0.27	0.33
	CAG	**1.64**	1.73	**1.93**	**1.81**	**1.69**	**1.57**	**1.73**	**1.67**
Arn (N)	AAU	0.87	0.99	0.96	0.89	0.89	1.1	0.55	0
	AAC	1.13	1.01	1.04	1.11	1.11	0.9	1.45	2
Lys (K)	AAA	0.23	0.8	**1.3**	0.78	0.72	0.63	0.48	**1.14**
	AAG	**1.77**	**1.2**	0.7	**1.22**	**1.28**	**1.38**	**1.52**	0.86
Asp (D)	GAU	**0.98**	**1.03**	**1.34**	**1.3**	**1.17**	0.89	**1.13**	**1**
	GAC	**1.02**	**0.97**	0.66	0.7	0.83	**1.11**	0.87	**1**
Glu (E)	GAA	0.37	0.38	0.67	0.36	0.36	0.38	0.52	0.4
	GAG	**1.63**	**1.62**	**1.33**	**1.64**	**1.64**	**1.62**	**1.48**	**1.6**
Cys (C)	UGU	0.29	0.67	0.43	0.53	0.59	0.4	0.64	1.2
	UGC	**1.71**	1.33	**1.57**	**1.47**	1.41	**1.6**	1.36	0.8

Shafat *et al. Journal of Genetic Engineering and Biotechnology*   (2022) 20:56

Page 7 of 14

Table 2 (continued)

AA	Codon	HEV	Human	Rabbit	Mongoose	Pig	Wild boar	Camel	Monkey
Arg (R)	CGU	**1.88**	**2.16**	**1.92**	**2.12**	**1.92**	**1.73**	**2.19**	**1.41**
	CGC	**1.65**	**1.35**	**2.47**	**1.76**	1.7	**1.81**	1.27	**2.82**
	CGA	0.3	0.34	0.24	0.35	0.26	0.3	0.42	0.35
	CGG	**1.35**	1.22	0.67	**1.29**	1.1	1.18	1.13	0.71
	AGA	0.34	0.3	0	0.35	0.31	0.3	0.21	0
	AGG	0.48	0.64	0.71	0.12	0.7	0.68	0.78	0.71
Gly (G)	GGU	**1.24**	**1.4**	**1.65**	**1.52**	**1.13**	0.92	**1.8**	0.86
	GGC	**1.79**	**1.8**	**2.32**	**1.81**	**1.91**	**2.11**	**1.45**	**2.29**
	GGA	0.23	0.19	0.03	0.57	0.26	0.18	0.35	0.29
	GGG	0.74	0.62	0	0.1	0.7	0.79	0.41	0.57

The preferred codons are indicated in bold

Fig. 2 Comparative analysis of relative synonymous codon usage (RSCU) patterns between HEV and its hosts (human, rabbit, mongoose, pig, wild boar, camel, and monkey)

codons (CUU, CUC, GUU, GUC, UCU, CAG, CGU, GGU), while the remaining 16 preferred codons had RSCU values >0.6 and <1.6. Out of these 16, 2 codons had RSCU values <1, i.e., less-abundant codons (UUC, GCA), while 14 had RSCU values >1, i.e., abundant codons (UUU, AUU, AUC, CCU, ACU, ACC, GCU, GCC, UAU, CAU, AAG, GAU, GAG GGC). No optional synonymous codon was underrepresented (RSCU < 0.6).

Monkey
Among the 27 preferred codons, 18 preferred codons were U/C-ending (U-ending: 10; C-ending: 8) and 9 were G/A -ending (G-ending: 5; A-ending: 4) (Table 2) (Additional file 16: S16 Table). This result inferred that U- and C-ending are preferred in YDR. Within these preferred

codons, 13 had RSCU value >1.6, i.e., overrepresented codons (CUU, CUC, AUU, GUU, GAG, UCC, AGU, CCA, GCG, CAG, GAG, CGC, GGC), while the remaining 14 preferred codons had RSCU values >0.6 and <1.6. Out of these 14, 3 codons had RSCU values <1, i.e., less-abundant codons (AUC, UAC, GCU), while 9 had RSCU values >1, i.e., abundant codons (GUC, UCU, CCG, ACG, GCA, UAU, CAU, AAA, CGU). In addition to this, 2 had RSCU values 1, i.e., random codons (GAU, GAC). No optional synonymous codon was underrepresented (RSCU < 0.6).

In line with compositional analysis, the RSCU analysis confirmed the codon biasness towards U- and C-ended codons. The RSCU pattern clearly indicated that the selection of preferred codons showed common

attributes as well as differences among HEV and HEV-hosts (Table 2). It was observed that some of the codons showed similar preference among HEV and HEV-hosts, while for other codons, HEV showed preference differed from that of its hosts or vice-versa. Thus, the codon which is most common among HEV and HEV-hosts, is considered as the most preferred codon that codes for a particular amino acid. Because the optimal codon selection in viruses largely depends on their hosts, we next compared the codon usage frequency of HEV with its hosts by correlating their RSCU patterns.

Relationship among HEV-hosts by comparing codon usage frequency

Since a particular amino acid is encoded by a preferred codon, the usage of synonymous codons is not random. Thus, we calculated the frequency of the preferred codons for each amino acid using the RSCU analysis (Additional file 9: S9 Table, Additional file 10: S10 Table, Additional file 11: S11 Table, Additional file 12: S12 Table, Additional file 13: S13 Table, Additional file 14: S14 Table, Additional file 15: S15 Table and Additional file 16: S16 Table), to analyze the relationship among HEV and its hosts. This was done to understand the influence of selection pressure from hosts on codon usage patterns of HEV. A list of preferred codons encoding amino acids with higher frequency as compared to other synonymous codons for HEV, and all the hosts were computed and compared as mentioned in Table 3.

The observed 4 amino acids Phe, His, Gln, and Glu showed similar usage of preferred codons (UUU, CAU, CAG, and GAG) among HEV and its hosts, which implicates an evidence of mutual codon preference. While few amino acids also showed differences in their choice of preferred codons. HEV and other HEV-hosts (human, rabbit, mongoose, pig, wild boar, camel) shared evidence of preferred codons (GUC, UGC, and CGU) for encoding the amino acids Val, Cys, and Arg, respectively, except for monkey which used different set of preferred codons (GUU, UGU, and CGC). Moreover, this phenomenon was also observed in other hosts, i.e., preferred codons encoding amino acids was different in specific host in comparison to other HEV-hosts and HEV. Firstly, HEV and HEV-hosts (human, mongoose, pig, wild boar, camel, and monkey) shared evidence of preferred codon for CCU which encoded Pro, except for rabbit which preferred CCC over CCU. Secondly, HEV and HEV-hosts (human, mongoose, rabbit, pig, camel, and monkey) shared evidence of preferred codon for AAC for encoding Asn, except for wild boar, which preferred AAU over AAC. Thirdly, HEV and HEV-hosts (human, mongoose, rabbit, pig, wild boar, and monkey) shared evidence of preferred codon for GGC for encoding Gly, except for camel which preferred GGU over GGC (Table 3).

In detail, among the 18 preferred codons in HEV, 13 were common between HEV and human; 11 were common between HEV and rabbit; 15 were common between HEV and mongoose; 13 were common between HEV and pig; 13 were common between HEV and wild boar; 12 were common between HEV and camel; and 8 were common between HEV and monkey (Table 3). Therefore, the abovementioned codons were common between HEV and respective hosts, indicating coincident codon usage portion, i.e., these preferred codons were commonly shared between the virus and host. However, discrepancies were also observed within the preferred codons between HEV and its hosts, i.e., dissimilar usage of preferred codons. Thus, the ratio of coincident/antagonist preferred codons was 13/5 between HEV and human; 11/7 between HEV and rabbit; 15/3 between HEV and mongoose; 13/5 between HEV and pig; 13/5 between HEV and wild boar; 12/6 between HEV and camel; and 8/10 between HEV and monkey. Thus, codon usage pattern of HEV YDR is a mix of coincidence and antagonism with respect to its hosts.

Thus, for a particular amino acid, if a preferred codon in HEV showed similarity with its host cell, this phenomenon is termed as "mutual codon preference of host–pathogens". This implies that similar codon usage pattern among HEV and HEV-hosts could help the virus to synthesize the amino acid and corresponding proteins in a more efficient manner, thus helping the pathogen to thrive in its host cells. On the contrary, the difference in preferred codon among HEV and HEV-hosts suggests lack of shared codon preference, causing reduction in the translation efficiency of the corresponding amino acids.

A heat map was constructed using RSCU values of various HEV strains and its hosts (Fig. 3), which revealed that HEV codon preference neither completely differed nor completely showed similarity with its hosts, indicating a mixture of similar and dissimilar codon preferences (Fig. 3). Moreover, the top five most and least frequent used codons were also identified which showed common attributes and differences in codon usage patterns of HEV isolates (Table 4).

Thus, our results clearly indicated that the synonymous codon usage of HEV is a mixture of the two types of codon usage: "coincidence and antagonism."

Effect of natural selection in shaping codon usage patterns

It has been suggested that the frequencies of nucleotides A and U /T should be equal to that of C and G at the third position of the codon if mutational pressure affects the synonymous codon usage bias [28]. However, huge variations were noted in the nucleotide base composition in case of all the hosts, signifying that synonymous codon usage bias could majorly be influenced by natural

Shafat *et al. Journal of Genetic Engineering and Biotechnology*   (2022) 20:56

Page 9 of 14

Table 3 Preferred codons for each amino acid in the YDR of HEV and its hosts

	HEV	Human	Rabbit	Mongoose	Pig	Wild boar	Camel	Monkey
Phe								
UUU(F)	5.2	6.6	7	7	6.5	5.8	5	6
UUC(F)	3.8	2.4	3.8	2	2.5	3.2	4	3
Leu								
UUA(L)	1.3	0.9	0.6	1.3	0.9	0.2	1.8	1
UUG(L)	2.6	2.5	0.8	2.7	2.1	2.5	2.2	3
CUU(L)	5.5	6	8	4	6.9	5.7	5.4	6
CUC(L)	6.1	5.4	6	6.3	5.2	5.7	6	5
CUA(L)	1.2	1	0.2	0.3	1.4	1.1	0	1
CUG(L)	2.6	3.7	4	4.3	3	4	3.4	2
Iso								
AUU(I)	4.6	4.6	5.8	2.7	5.1	4.3	5.4	8
AUC(I)	4.8	4.5	4.9	5.3	4.2	4.1	5.2	4
AUA(I)	2.2	3.3	2.1	5	3.4	4	2	1
Val								
GUU(V)	5.5	5.3	4.2	4.7	5.3	4.5	7.6	11
GUC(V)	6.4	5.3	6.7	7	5.3	6.4	7.8	6
GUA(V)	0.4	1	0.9	0	0.9	0.5	0.2	0
GUG(V)	3.8	4.8	4.3	5.3	5	5.5	1	0
Ser								
UCU(S)	3.2	4.7	4.6	5.7	4.5	4.7	5	4
UCC(S)	5.8	4	2.6	3	3.5	3.8	2.6	5
UCA(S)	2.2	2.4	2.4	2.7	2.9	1.5	1.6	2
UCG(S)	2	2.1	3.3	1.7	2.3	3.1	2.6	1
AGU(S)	1.6	1.4	1	1.3	1.6	2.4	2.4	3
AGC(S)	1.4	1.6	2	1.7	1.5	0.8	2	2
Pro								
CCU(P)	5	4.3	2.8	5.7	4.4	3.6	4.4	7
CCC(P)	2.4	2.3	4.7	2	2	2.3	3	2
CCA(P)	2.6	2.6	1.7	1.3	2.9	2.9	2	2
CCG(P)	2.1	2.7	3	3	2.8	3.1	2.4	1
Thr								
ACU(T)	5.3	6.2	5.6	7.3	5.5	5.8	6.4	8
ACC(T)	8.8	5.2	6.2	4.7	5.5	5.6	5.6	5
ACA(T)	2.5	3.5	3.2	1.7	3.4	3.6	3.2	3
ACG(T)	2.2	1.6	1.8	2.3	1.7	1.4	1.6	0
Ala								
GCU(A)	5.9	4.8	6.8	5.7	4.6	6.7	6.2	8
GCC(A)	9	7.9	6.8	6.7	8.1	6.4	7	7
GCA(A)	1.9	3.4	1.6	3	4	3.5	3.8	4
GCG(A)	0.9	3.7	4.4	4.7	3.2	3.3	3.4	1
Tyr								
UAU(Y)	5.5	5.9	5.2	5.3	6.2	6.4	8.8	7
UAC(Y)	6.5	6	5	6.7	5.8	5.6	3.2	5
His								
CAU(H)	4.2	4.7	4.1	4.3	4.7	5.1	4.4	4
CAC(H)	2.8	2.3	2.9	2.7	2.3	1.8	2.6	3
Gln								
CAA(Q)	1.1	0.9	0.2	0.7	1.1	1.5	0.8	1
CAG(Q)	5.2	6	6.3	6.3	5.9	5.5	5.2	5
Asn								
AAU(N)	1.6	1.5	1.4	1.3	1.3	1.7	0.6	0
AAC(N)	2.1	1.6	1.6	1.7	1.6	1.4	1.6	2
Lys								
AAA(K)	0.7	2.6	3.9	2.3	2.3	1.9	1.6	4
AAG(K)	5.5	3.9	2.1	3.7	4.1	4.2	5	3
Asp								
GAU(D)	3.9	4.1	5.2	5	4.6	3.5	4.4	4
GAC(D)	4.1	3.9	2.6	2.7	3.3	4.3	3.4	4
Glu								
GAA(E)	1.8	1.7	3.2	1.7	1.6	1.7	2.6	2
GAG(E)	7.9	7.3	6.4	7.7	7.5	7.3	7.4	8
Cys								
UGU(C)	0.7	1.7	1.1	1.3	1.5	1	1.6	3
UGC(C)	4.3	3.3	4.1	3.7	3.5	4	3.4	2
Arg								
CGU(R)	5.5	6	5.4	6	5.3	4.8	6.2	4
CGC(R)	4.8	3.7	7	5	4.7	5.1	3.6	8
CGA(R)	0.9	1	0.7	1	0.7	0.8	1.2	1
CGG(R)	3.9	3.4	1.9	3.7	3	3.3	3.2	2
AGA(R)	1	0.8	0	1	0.9	0.8	0.6	0
AGG(R)	1.4	1.8	2	0.3	2	1.9	2.2	2
Gly								
GGU(G)	4.3	4.9	5.8	5.3	4	3.2	6.2	3
GGC(G)	6.3	6.3	8.1	6.3	6.7	7.4	5	8
GGA(G)	0.8	0.7	0.1	2	0.9	0.6	1.2	1
GGG(G)	2.6	2.2	0	0.3	2.5	2.8	1.4	2

Table 3 (continued)

Comparison of codon usage frequency of preferred codons among HEV and its hosts. All the preferred codons are highlighted indicating the highest codon frequency. Thus, codon usage pattern of YDR was a mix of coincidence and antagonism with respect to its host

selection (Table 1). From these findings, it was clear that compositional constraints under mutation pressure combined with natural selection shaped the HEV YDR across all its hosts.

Discussion

Inspection of factors governing protein evolution is essential for various research fields, including comparative genomics, molecular evolution, and structural biology. With this study, we implemented a systematic survey of the evolutionary pressures (i.e., mutational bias and natural selection) across the YDR to gain insights into the HEV functional implications in regulation as well as adaptative evolution.

Jenkins and Holmes (2003) reported that codon usage bias phenomenon can be influenced by the overall nucleotide composition pattern [37]. Thus, initially, we computed the nucleotide frequencies of the YDR from HEV and its hosts. The HEV YDR revealed an over-representation of C, with overall C/U codon bias pattern in the nucleotide composition. In HEV, the percentage of C was the highest followed by U and G, with A having the lowest value (except hosts, camel, and monkey which followed the trend U > C > G > A). This clearly revealed that there was unequal distribution of A, U, G, and C nucleotides among the YDR codons. Additionally, in HEV and rabbit, the nucleotide values at third codon positions also followed the same trend, i.e., C3 had the highest value, followed by U3, G3, and A3 with the least value (while hosts followed the trend C3 > U3 > G3 > A3). Therefore, it could be interpreted that the initial nucleotide compositional patterns showed more preference towards C- and U-ended codons followed by G/A-ended codons. This is consistent with the recent investigation that has reported U/C rich genome in ORF1 of HEV [45]. However, the overall C/U rich pattern in the nucleotide content in YDR is opposite to the pattern observed in RNA viruses, which showed the prevalence of A/C-rich genomes (HIV, hepatitis C, rubella viruses) [46]. Thus, it could be interpreted that this biasness in YDR was due to the adaptation of common ancestor of modern HEV strains in terms of nucleotide composition requirement of the host during its process of evolution [47].

It has been suggested that particularly in viruses, AU- or GC-rich genomes tends to correlate with the RSCU patterns. For instance, AU- or GC-rich composition preferred codons ending with either A and U or G and C,
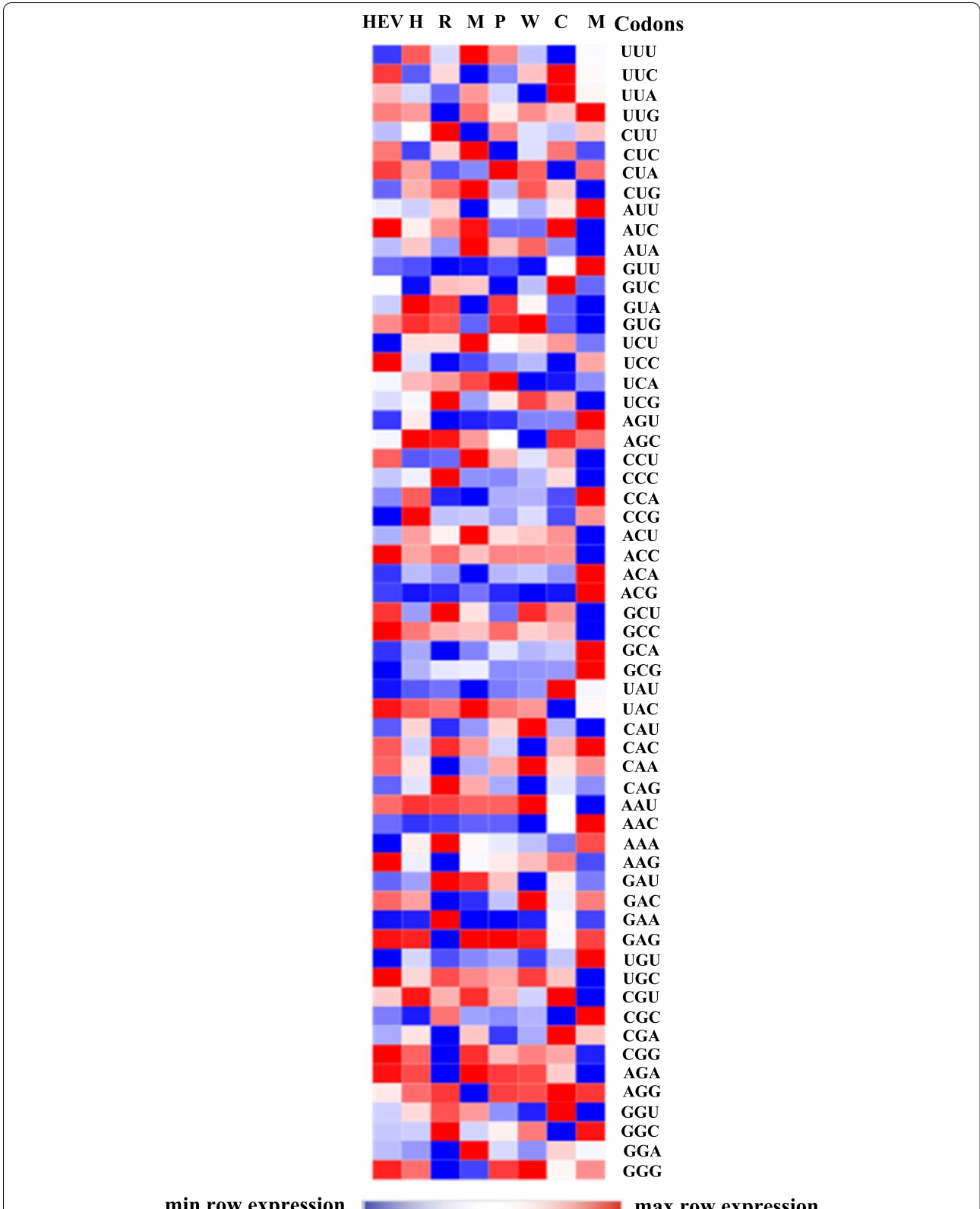
Fig. 3 Heat map showing the relative synonymous codon usage (RSCU) values accompanying different hosts (H: human, R: rabbit, M: mongoose, P: pig, W: wild boar, C: camel, and M: monkey). The host species are mentioned on the horizontal axis and codons are represented on the vertical axis. Heatmap confirms the occurrence of resemblance as well as discrepancies in RSCU pattern among different hosts

Shafat *et al. Journal of Genetic Engineering and Biotechnology* (2022) 20:56

Page 11 of 14

Table 4 Most frequent and least used codons among HEV and its natural hosts

Top 5 most frequent used codons	
HEV	ACC (8.8), GAG (7.9), UAC (6.5), GUC (6.4), GGC (6.3),
Human	GCC (7.9), UUU (6.6), GGC (6.3), ACU (6.2), CGU (6.0)
Rabbit	GGC (8.1), CUU (8.0), CGC/UUU (7.0), GCU/GCC (6.8), GUC (6.7)
Mongoose	GAG (7.7), ACU (7.3), UUU/GUC (7.0), GCC/UAC (6.7), GGC/CAG/CUC (6.3)
Pig	GCC (8.1), GAG (7.5), CUU (6.9), GGC (6.7), UUU (6.5)
WB	GGC (7.4), GAG (7.3), GCU (6.7), GUC/GCC/UAU (6.4), UUU/ACU (5.8)
Camel	UAU (8.8), GUC (7.8), GUU (7.6), GAG (7.4), GCC (7.0)
Monkey	GUU (11), AUU/ACU/GCU/GAG/CGC/GGC (8), CCU/GCC/UAU (7), UUU/CUU/GUC (6), CUC/UCC/ACC/UAC/CAG (5)
Top 5 least frequent used codons	
HEV	GUA (0.4), UGU/AAA (0.7), GGA (0.8), CGA/GCG (0.9), AGA (1)
Human	GGA (0.7), AGA (0.8), CAA/UUA (0.9), CUA/GUA/CGA (1.0), AGU (1.4)
Rabbit	AGA/GGG (0), GGA (0.1), CUA/CAA (0.2), UUA (0.6), CGA (0.7), UUG (0.8)
Mongoose	GUA (0), CUA/GGG/AGG (0.3), CAA (0.7), AGA/CGA (1.0), UUA/CCA/AGU/UGU/AAU (1.3), UCG/ACA/AAC/GAA/AGC (1.7),
Pig	CGA (0.7), UUA/GUA/AGA/GGA (0.9), CAA (1.1), AAU (1.3), CUA (1.4)
Wild boar	UUA (0.2), GUA (0.5), GGA (0.6), CGA/AGC/AGA (0.8), UGU (1.0)
Camel	CUA (0), GUA (0.2), AAU/AGA (0.6), CAA (0.8), GUG (1.0), CGA/GGA (1.2)
Monkey	GUA/GUG/ACG/AAU/AGA (0), UUA/CUA/AUAUCG/CCG/GCG/CAA/CGA/GGA (1), CUG/UCA/CCC/CCA/AAC/GAA/UGC/CGG/AGC/AGG/GGG (2), UUC/UUG/ACA/CAC/AAG/UGU/AGU/GGU (3), AUC/UCU/GCA/CAU/AAA/GAU/GAC/CGU (4), CUC/UCC/ACC/UAC/CAG (5)

Codon frequency is given in parentheses following the relative synonymous codon usage

respectively. These trends, when observed, support the influence of mutational pressure [37]. The RSCU analysis revealed that HEV had comparatively higher codon usage bias towards U- and C-ended codons. The overall RSCU patterns can potentially hide host-specific patterns, so we next calculated the RSCU values for specific hosts. Thus, the comparative analysis was performed among HEV and its hosts, by correlating their RSCU patterns. It was noted that the host-specific codon usage patterns also showed preferred codons ending with U and C. Thus, in line with nucleotide composition analysis, the RSCU analysis further confirmed the codon biasness towards U- and C-ended codons. Thus, it could be interpreted that mutational bias was found to be a major force determining the codon usage patterns of YDR, which probably suggested that compositional constraints influenced the selection of preferred codons. However, it is interesting to mention that though HEV and its hosts was endowed with higher percentage of GC rather than AC, the RSCU analysis revealed a biasness towards U-terminated codons. This suggested that other factors in combination with mutation pressure also existed in the process of HEV evolution. Therefore, selection pressure from hosts contributed to shaping the molecular evolution of HEV at the level of codon usage.

The codon usage in virus's genome in accordance with its host codon preferences is an important aspect which determines the evolutionary adaptation of the virus to its host cell. The alteration of codon usage in viral genomes due to the proper information obtained from host genes regulates the virus-host interactions [48]. As viruses are obligate parasites, their optimal codon selection is largely dependent on their host cells translational machinery [49]. A noteworthy variation was observed in the usage for the preferred codons among HEV and HEV-hosts. This implied that the codon usage patterns of HEV as well as the possible fitness of HEV to adapt within its dynamic host range were largely influenced by the selection pressures exerted from HEV-hosts.

In this study, it was observed, that unlike other viruses, that have evolved completely identical to their hosts or completely opposite to their hosts codon patterns [50, 51], the HEV evolution showed a mixture of two codon usage patterns. Our results revealed that none of the hosts showed complete resemblance or complete discrepancy to the HEV. The ratio of common/uncommon preferred codons between HEV-Human, HEV-Rabbit, HEV-Mongoose, HEV-Pig, HEV-Wild boar, HEV-Camel, and HEV-Monkey were 13/5, 11/7, 15/3, 13/5, 13/5, 12/6, and 8/10, respectively. Thus, codon usage pattern of HEV YDR showed a mixture of coincidence and antagonism with respect to its hosts. The resemblance in synonymous codon patterns among HEV and its hosts implied that HEV could adapt to its host cells, resulting in its multiplication. This phenomenon suggests that the virus can replicate in host cells due to similarity in usage for preferred

Shafat *et al. Journal of Genetic Engineering and Biotechnology* (2022) 20:56

Page 12 of 14

codons. It has been suggested that the coincident portions of codon usage could facilitate efficient translation of the corresponding amino acids among viruses and their respective hosts [52]. This indicates that preferred codons or more abundant tRNA molecules are chosen to increase the accuracy of translation [53]. While the antagonistic portions of codon usage may aid in proper folding of viral proteins, even though decrease in the corresponding amino acids translation efficiency is observed [52]. This implies that rare codons help in reducing the inappropriate co-translational folding of proteins [54]. Thus, our results clearly suggested that the codon usage pattern of HEV is both coincident and antagonistic to that of its hosts. Such patterns of coincidence and antagonism have been previously reported in HBV [55], HCV [52] and enterovirus 71 [29]. Therefore, our results probably suggested that disfavored codons encoding amino acids cannot be considered as a deleterious factor for viral genes in order to adapt to its hosts.

Thus, it could be interpreted that the influence of compositional constraints, codon usage biasness, and mutational alongside the selective forces were reflected in the occurrence of HEV YDR codon usage patterns.

Conclusions

To the best of knowledge, this report documents the codon usage analysis in HEV YDR for the first time using bioinformatics approach. Thisnovel approach is expected to strengthen our understanding on the common attributes and differences in the codon usage patterns among HEV and its various hosts. The nucleotide compositional analysis showed relative abundance of C and U nucleotides and relative synonymous codon usage analysis revealed that the preferred synonymous codons mostly end with C/U. Moreover, it was observed that the HEV codon usage pattern to that of its host cells is a mixture of coincidence and antagonism. The compositional characteristics indicated that interaction between the mutation pressure from virus and translation selection from host exist in the processes of HEV evolution. Our study suggested that synonymous codon usage in HEV is an evolutionary process, perhaps reflecting a dynamic process of mutation and selection forces to adjust its codon usage to different hosts and conditions. The present study is thus envisaged to infer the evolution, adaptation, and biology of HEV via specific codon preferences.

Abbreviations
HEV: Hepatitis E virus; YDR: Y-domain region; RSCU: Relative synonymous codon usage.

Supplementary Information

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Additional file 5.

Additional file 6.

Additional file 7.

Additional file 8.

Additional file 9.

Additional file 10.

Additional file 11.

Additional file 12.

Additional file 13.

Additional file 14.

Additional file 15.

Additional file 16.

Authors' contributions
SP conceptualized the research. SP and ZS designed the manuscript. ZS was a major contributor in writing the manuscript and performed the biocomputational analysis of the protein. KP and AA proofread the manuscript. The authors read and approved the final manuscript.

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India. [2]Centre of Excellence in Biotechnology Research, College of Science, King Saud University, Riyadh, Saudi Arabia. [3]Department of Pharmacognosy, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia.

Shafat *et al. Journal of Genetic Engineering and Biotechnology*   (2022) 20:56

Page 13 of 14

References

1. Lhomme S, Marion O, Abravanel F, Chapuy-Regaud S, Kamar N, Izopet J (2016) Hepatitis E pathogenesis. Viruses 8(8):212
2. Kamar N, Izopet J, Pavio N, Aggarwal R, Labrique A, Wedemeyer H, Dalton HR (2017) Hepatitis E virus infection. Nat Rev Dis Primers 3(1):1–6
3. Tam AW, Smith MM, Guerra ME, Huang CC, Bradley DW, Fry KE, Reyes GR (1991) Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. Virology 185(1):120–131
4. Ansari IH, Nanda SK, Durgapal H, Agrawal S, Mohanty SK, Gupta D, Jameel S, Panda SK (2000) Cloning, sequencing, and expression of the hepatitis E virus (HEV) nonstructural open reading frame 1 (ORF1). J Med Virol 60(3):275–283
5. Parvez MK (2013) Molecular characterization of hepatitis E virus ORF1 gene supports apapain-like cysteine protease (PCP)- domain activity. Virus Res 178(2):553–556
6. Chandra V, Taneja S, Kalia M, Jameel S (2008) Molecular biology and pathogenesis of hepatitis E virus. J Biosci 33(4):451–464
7. Mori Y, Matsuura Y (2011) Structure of hepatitis E viral particle. Virus Res 161(1):59–64
8. He M, Wang M, Huang Y, Peng W, Zheng Z, Xia N, Xu J, Tian D (2016) The ORF3 protein of genotype 1 hepatitis E virus suppresses TLR3-induced NF-kB signaling via TRADD and RIP1. Sci Rep 6(1):1–13
9. Parvez MK, Al-Dosari MS (2015) Evidence of MAPK-JNK1/2 activation by hepatitis E virus ORF3 protein in cultured hepatoma cells. Cytotechnology 67(3):545–550
10. Ding Q, Heller B, Capuccino JM, Song B, Nimgaonkar I, Hrebikova G, Contreras JE, Ploss A (2017) Hepatitis E virus ORF3 is a functional ion channel required for release of infectious particles. Proc Natl Acad Sci USA 114(5):1147–1152
11. Parvez MK (2017) Mutational analysis of hepatitis E virus ORF1 "Y-domain": effects on RNA replication and virion infectivity. World J Gastroenterol 23(4):590
12. Wu J, Si F, Jiang C, Li T, Jin M (2015) Molecular detection of hepatitis E virus in sheep from southern Xinjiang, China. Virus Genes 50(3):410–417
13. Meng XJ (2011) From barnyard to food table: the omnipresence of hepatitis E virus and risk for zoonotic infection and food safety. Virus Res 161(1):23–30
14. Takahashi K, Terada S, Kokuryu H, Arai M, Mishiro S (2010) A wild boar-derived hepatitis E virus isolate presumably representing so far unidentified "genotype 5". Kanzo 51(9):536–538
15. Takahashi M, Nishizawa T, Sato H, Sato Y, Nagashima S, Okamoto H (2011) Analysis of the full-length genome of a hepatitis E virus isolate obtained from a wild boar in Japan that is classifiable into a novel genotype. J Gen Virol 92(4):902–908
16. Sarchese V, Di Profio F, Melegari I, Palombieri A, Sanchez SB, Arbuatti A, Ciuffetelli M, Marsilio F, Martella V, Di Martino B (2019) Hepatitis E virus in sheep in Italy. Transbound Emerg Dis 66(3):1120–1125
17. Tei S, Kitajima N, Takahashi K, Mishiro S (2003) Zoonotic transmission of hepatitis E virus from deer to human beings. Lancet 362:371–373
18. Takahashi K, Kitajima N, Abe N, Mishiro S (2004) Complete or near-complete nucleotide sequences of hepatitis E virus genome recovered from a wild boar, a deer, and four patients who ate the deer. Virology 330:501–505
19. Nakamura M, Takahashi K, Taira K, Taira M, Ohno A, Sakugawa H, Arai M, Mishiro S (2006) Hepatitis E virus infection in wild mongooses of Okinawa, Japan: demonstration of anti-HEV antibodies and a full-genome nucleotide sequence. Hepatol Res 34:137–140
20. Xu F, Pan Y, Baloch AR, Tian L, Wang M, Na W, Ding L, Zeng Q (2014) Hepatitis E virus genotype 4 in yak, northwestern China. Emerg Infect Dis 20(12):2182
21. Yugo DM, Cossaboom CM, Heffron CL, Huang YW, Kenney SP, Woolums AR, Hurley DJ, Opriessnig T, Li L, Delwart E, Kanevsky I (2019) Evidence for an unknown agent antigenically related to the hepatitis E virus in dairy cows in the United States. J Med Virol 91(4):677–686
22. Sanford BJ, Emerson SU, Purcell RH, Engle RE, Dryman BA, Cecere TE, Buechner-Maxwell V, Sponenberg DP, Meng XJ (2013) Serological evidence for a hepatitis E virus (HEV)-related agent in goats in the United States. Transbound Emerg Dis 60(6):538–545
23. Izopet J, Dubois M, Bertagnoli S, Lhomme S, Marchandeau S, Boucher S, Kamar N, Abravanel F, Guérin JL (2012) Hepatitis E virus strains in rabbits

and evidence of a closely related strain in humans, France. Emerg Infect Dis 18(8):1274
24. Rasche A, Saqib M, Liljander AM, Bornstein S, Zohaib A, Renneker S, Steinhagen K, Wernery R, Younan M, Gluecks I, Hilali M (2016) Hepatitis E virus infection in dromedaries, North and East Africa, United Arab Emirates, and Pakistan, 1983–2015. Emerg Infect Dis 22(7):1249
25. Woo PCY, Lau SKP, Teng JLL, Cao KY, Wernery U, Schountz T, Chiu TH, Tsang AKL, Wong PC, Wong EYM et al (2016) New hepatitis E virus genotype in Bactrian camels, Xinjiang, China, 2013. Emerg Infect Dis 22:2219–2221
26. Grantham R, Gautier C, Gouy M, Mercier R, Pave A (1980) Codon catalog usage and the genome hypothesis. Nucleic Acids Res 8(1):197–197
27. Marin A, Bertranpetit J, Oliver JL, Medina JR (1989) Variation in G+C-content and codon choice: differences among synonymous codon groups in vertebrate genes. Nucleic Acids Res 17(15):6181–6189
28. Gu W, Zhou T, Ma J, Sun X, Lu Z (2004) Analysis of synonymous codon usage in SARS coronavirus and other viruses in the Nidovirales. Virus Res 101(2):155–161
29. Liu YS, Zhou JH, Chen HT, Ma LN, Pejsak Z et al (2011) The characteristics of the synonymous codon usage in enterovirus 71 virus and the effects of host on the virus in codon usage pattern. Infect Genet Evol 11(5):1168–1173
30. Ma JJ, Zhao F, Zhang J, Zhou JH, Ma LN et al (2013) Analysis of synonymous codon usage in dengue viruses. J Anim Vet Adv 12(1):88–98
31. Moratorio G, Iriarte A, Moreno P, Musto H, Cristina J (2013) A detailed comparative analysis on the overall codon usage patterns in West Nile virus. Infect Genet Evol 14:396–400
32. Sharp PM, Cowe E, Higgins DG, Shields DC, Wolfe KH et al (1988) Codon usage patterns in Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens; a review of the considerable within-species diversity. Nucleic Acids Res 16(17):8207–8211
33. Tao P, Dai L, Luo M, Tang F, Tien P et al (2009) Analysis of synonymous codon usage in classical swine fever virus. Virus Genes 38(1):104–112
34. Sharp PM, Li WH (1986) Codon usage in regulatory genes in Escherichia coli does not reflect selection for 'rare' codons. Nucleic Acids Res 14(19):7737–7749
35. Duret L, Mouchiroud D (1999) Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. Proc Natl Acad Sci U S A 96(8):4482–4487
36. Van der Linden MG, de Farias ST (2006) Correlation between codon usage and thermostability. Extremophiles 10(5):479–481
37. Jenkins GM, Holmes EC (2003) The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res 92(1):1–7
38. Wang M, Zhang J, Zhou JH, Chen HT, Ma LN et al (2011) Analysis of codon usage in bovine viral diarrhea virus. Arch Virol 156(1):153–160
39. Wong EH, Smith DK, Rabadan R, Peiris M, Poon LL (2010) Codon usage bias and the evolution of influenza A viruses. Codon usage biases of influenza virus. BMC Evol Biol 10(1):253
40. Chen Y (2013) A comparison of synonymous codon usage bias patterns in DNA and RNA virus genomes: quantifying the relative importance of mutational pressure and natural selection. Biomed Res Int 2013:406342
41. Shi SL, Jiang YR, Liu YQ, Xia RX, Qin L (2013) Selective pressure dominates the synonymous codon usage in parvoviridae. Virus Genes 46(1):10–19
42. Zhang Z, Dai W, Wang Y, Lu C, Fan H (2013) Analysis of synonymous codon usage patterns in torque teno sus virus 1 (TTSuV1). Arch Virol 158(1):145–154
43. Zhang Z, Dai W, Dai D (2013) Synonymous codon usage in TTSuV2: analysis and comparison with TTSuV1. PLoS One 8:e81469
44. Shackelton LA, Parrish CR, Holmes EC (2006) Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J Mol Evol 62(5):551–563
45. Baha S, Behloul N, Liu Z, Wei W, Shi R, Meng J (2019) Comprehensive analysis of genetic and evolutionary features of the hepatitis E virus. BMC Genomics 20(1):1–16
46. Auewarakul P (2005) Composition bias and genome polarity of RNA viruses. Virus Res 109(1):33–37
47. Bouquet J, Cherel P, Pavio N (2012) Genetic characterization and codon usage bias of full-length hepatitis E virus sequences shed new lights on

Shafat *et al. Journal of Genetic Engineering and Biotechnology*   (2022) 20:56

Page 14 of 14

genotypic distribution, host restriction and genome evolution. Infect Genet Evol 12(8):1842–1853

48. Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S (2008) Virus attenuation by genome-scale changes in codon pair bias. Science 320(5884):1784–1787

49. Zhou H, Wang H, Huang LF, Naylor M, Clifford P (2005) Heterogeneity in codon usages of sobemovirus genes. Arch Virol 150:1591–1605

50. Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E (2006) Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. J Virol 80:9687–9696

51. Sanchez G, Bosch A, Pinto RM (2003) Genome variability and capsid structural constraints of hepatitis a virus. J Virol 77:452–459

52. Hu JS, Wang QQ, Zhang J, Chen HT, Xu ZW, Zhu L, Ding YZ, Ma LN, Xu K, Gu YX, Liu YS (2011) The characteristic of codon usage pattern and its evolution of hepatitis C virus. Infect Genet Evol 11:2098–2102

53. Hershberg R, Petrov DA (2008) Selection on codon bias. Annu Rev Genet 42:287–299

54. Komar AA, Lesnik T, Reiss C (1999) Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. FEBS Lett 462(3):387–391

55. Ma MR, Ha XQ, Ling H, Wang ML, Zhang FX, Li G, Yan W (2011) The characteristics of the synonymous codon usage in hepatitis B virus and the effects of host on the virus in codon usage pattern. Virology J 8(1):1–0

Publisher's Note