Research paper

Analysis of codon usage bias of Crimean-Congo hemorrhagic fever virus and its adaptation to hosts

Siddiq Ur Rahman[a,1], Xiaoting Yao[a,1], Xiangchen Li[a], Dekun Chen[b,*], Shiheng Tao[a,*]

[a] College of Life Sciences and State Key Laboratory of Crop Stress Biology in Arid Areas, Northwest A&F University, Yangling, Shaanxi, China
[b] College of Veterinary Medicine and State Key Laboratory of Crop Stress Biology in Arid Areas, Northwest A&F University, Yangling, Shaanxi, China

ARTICLE INFO

ABSTRACT

Crimean-Congo hemorrhagic fever virus (CCHFV) is a negative-sense, single stranded RNA virus with a three-segmented genome that belongs to the genus *Nairovirus* within the family *Bunyaviridae*. CCHFV uses *Hyalomma* ticks as a vector to infect humans with a wide range of clinical signs, from asymptomatic to Zika-like syndrome. Despite significant progress in genomic analyses, the influences of viral relationships with different hosts on overall viral fitness, survival, and evading the host's immune systems remain unknown. To better understand the evolutionary characteristics of CCHFV, we performed a comprehensive analysis of the codon usage pattern in 179 CCHFV strains by calculating the relative synonymous codon usage (RSCU), effective number of codons (ENC), codon adaptation index (CAI), and other indicators. The results indicate that the codon usage bias of CCHFV is relatively low. Several lines of evidence support the hypothesis that a translation selection factor is shaping codon usage pattern in this virus. A correspondence analysis (CA) showed that other factors, such as base composition, aromaticity, and hydrophobicity may also be involved in shaping the codon usage pattern of CCHFV. Additionally, the results from a comparative analysis of RSCU between CCHFV and its hosts suggest that CCHFV tends to evolve codon usage patterns that are comparable to those of its hosts. Furthermore, the selection pressures from *Homo sapiens*, *Bos taurus*, and *Ovis aries* on the CCHFV RSCU patterns were dominant when compared with selection pressure from *Hyalomma* spp. vectors. Taken together, both natural selection and mutation pressure are important for shaping the codon usage pattern of CCHFV. We believe that such findings will assist researchers in understanding the evolution of CCHFV and its adaptation to its hosts.

1. Introduction

Crimean-Congo hemorrhagic fever (CCHFV) is a negative-sense, single stranded RNA virus that belongs to the *Bunyaviridae* family. The virus has a segmented genome specialized for different functions: S (small), M (medium), and L (large) segments encode nucleocapsid (N), glycoproteins (Gn and Gc), and RNA-dependent polymerases (RdRp), respectively (Bente et al., 2013). The M segment of this genus (*Nairovirus*) is comparatively larger than others in the family and codes for a protein of around 240 kDa (Elliott, 2017; Papa et al., 2002). The genetic diversity of the M segment in the CCHFV genome enhances the variation frequency of glycoproteins (Gn and Gc) in the viral life cycle. Glycoproteins bind cell receptor recognition sites and facilitate viral infection in different vertebrate hosts (Bente et al., 2013; Bertolotti-ciarlet et al., 2005; Peyrefitte et al., 2015). Like other members of the *Bunyaviridae* family, the 3′ and 5′ terminal sequences of each genome segment are conserved and complementary to each other, forming a pan-handle structure that possesses conserved polymerase binding sites (Peyrefitte et al., 2015). Due to their roles in immunity, pathogenicity, and vaccine development, all three genome segments were included in the current study.

CCHFV causes a tick-borne zoonotic infection (Crimean-Congo hemorrhagic fever (CCHF)) and is among the deadliest human pathogens in Africa and Eurasia (Shayan et al., 2015). CCHFV infection is transmitted to humans and animals through its main vector, ticks of the genus *Hyalomma* (Ixodidae). The same vector is also responsible for spreading *Theileria*, Kyasanur Forest disease virus (KFDV), and *Babesia* (Ghosh and Nagar, 2014). Specific tick-host cycles therefore have a strong influence on the circulation of CCHFV in its natural foci. Some species wait passively to encounter a vertebrate host ("ambush ticks"), but *Hyalomma* species, also called hunting ticks, have the ability to hunt up to a distance of 400 m in order to find their hosts (including humans) (Bente et al., 2013).

In the past, CCHFV infection was characterized by non-specific signs

and symptoms like high fever, headache, stomach pain, myalgia, joint pain, and vomiting which were similar to other tropical infections, caused by highly pathogenic viruses, such as Ebola virus (EBOV) and Marburg virus (MARV) (Nasrullah et al., 2015). However, in severe cases CCHFV may cause hemorrhage, with a high case-fatality rate (European Centre for Disease Prevention and Control, 2016; Leblebicioglu et al., 2016; World Health Organization, 2013). CCHFV does not show any obvious signs of illness in host animals other than humans, and thus, the information of virus distribution in a given geographic location is based on the incidence of disease in humans (Bente et al., 2013).

CCHF is a fatal viral disease, with a long history in human populations. CCHFV infection was identified for the first time in 1944, in a Crimean region where 10% of Soviet troops were infected, followed by further spread to Bulgaria and South Africa (Aslam et al., 2016). From 1967 until late 2000, many cases of CCHFV infection were identified in south Asia and central Africa, and infections spread into southeastern Europe, creating a world health emergency (Bente et al., 2013). The incidence of the disease has clearly been increasing; from 2002 to 2008, there were > 1000 confirmed cases, with a 3.2% mortality rate (Bente et al., 2013; Butenko and Karganova, 2007; Leblebicioglu, 2010). Genetic analyses indicated a close association between the central Africa strains and the Eurasian strains (Bente et al., 2013; Hoogstraal, 1979). Furthermore, phylogenetic analysis revealed greater genetic diversity in CCHFV than in any other arthropod-borne virus. This diversity is related to virus-infected ticks occurring on migratory birds, which travel throughout the world, coming into contact with livestock and surmounting topographic barriers (Bente et al., 2013; Chamberlain et al., 2005; Hewson et al., 2004).

Apart from transmission by ticks, other routes of transmission for CCHFV include infected mother to offspring, sexual contact, blood transfusion, and contact with different viremic body fluids of patients and infected animals (Metanat et al., 2016; Yurievna et al., 2016). The primary hosts of CCHFV include a wide range of domestic animals such as cattle, sheep, and goats, probably leading to increased population-based outbreaks (Leblebicioglu, 2010); humans, however, are dead-end hosts.

Synonymous codon usage bias has been studied in a wide range of organisms, from prokaryotes to eukaryotes and viruses (Butt et al., 2016). Generally, 61 triplets encode and arrange 20 different amino acids, therefore, many of them are synonymous in expression. Interestingly, synonymous codons are used in different frequencies by various organisms or even in different gene groups of the same genome; this phenomenon was termed codon usage bias (Hershberg and Petrov, 2008; Andersson and Kurland, 1990). Studies on codon usage highlighted several factors, ranging from protein translation to folding, that might influence patterns of codon usage: translation, mutational pressure, protein secondary motifs, replication, and transcriptional factors (Cristina et al., 2015; Rahman et al., 2017). Among these, compositional constraints under mutational pressure and natural selection are considered to be the major factors responsible for codon usage variation among different organisms (Butt et al., 2016).

Various studies of codon usage in different viruses have revealed mutational pressure as the major factor shaping codon usage patterns compared with natural selection (Cristina et al., 2015; Sharp et al., 2010). However, as our understanding of codon usage improves, it appears that although mutational pressure is a major driving force, it is certainly not the only one when different types of RNA and DNA viruses are considered (Butt et al., 2014). Considering their small genome size in comparison with prokaryotic and eukaryotic genomes, and features such as dependence on host machinery for key processes including replication, protein synthesis, and transmission, the interplay of codon usage between viruses and their hosts is expected to affect overall viral fitness, survival, and avoidance of host cell responses and evolution (Burns et al., 2006; Costafreda et al., 2014; Mueller et al., 2006). Thus, knowledge of codon usage in viruses provides information on molecular

evolution, and also improves our understanding of the regulation of viral gene expression and aids in vaccine design, where the efficient expression of viral proteins may be required to generate immunity.

This study focuses on 179 different strains of CCHFV related to African and Eurasian lineages. We performed genomic analyses for codon usage using available and complete S, M, and L segment sequence data. We analyzed evolutionary adaptation of CCHFV to its hosts and explored factors that play an important role in shaping codon usage patterns in the CCHFV genome.

2. Material and methods

2.1. Data collection

In this study, a total of 179 CCHFV complete genomic sequences representing S, M, and L segments were obtained from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/GenBank). Table S1 shows the accession numbers and strain names. For each strain, the ORFs were obtained by Lasergene SeqBuilder (Singh et al., 2016) and aligned using the MUSCLE program (Goñi et al., 2012).

2.2. Nucleotide contents analysis

Nucleotide compositional analysis of the 179 CCHFV coding sequences were measured using CodonW (http://sourceforge.net/projects/codonw, written by John Peden). The overall nucleotide occurrence frequencies of U, G, C, and A, percentage nucleotide occurrence at the 3rd codon position (U3, G3, C3, and A3), and G + C nucleotide occurrence at the 1st (GC1), 2nd (GC2), and 3rd (GC3) positions were calculated. In addition, the mean frequency of G + C at GC1-2 positions and the total AU/GC contents were also measured. Along with the three stop codons, AUG and UGG (no synonymous codons), were excluded in the current analysis.

2.3. Data analysis

To determine the characteristics for codon usage bias, the relative synonymous codon usage (RSCU) was calculated for all previously described sequences (Sharp and Li, 1986). The relative abundance of dinucleotide frequencies in the polyprotein-coding regions of CCHFV were computed with the SSE v1.2 editor (Karniychuk, 2016; Simmonds, 2012). For the complete synonymous codon usage of consecutive genes, total mean frequencies of GC content that occurred at the 1st, 2nd, and 3rd codon positions were obtained using the effective number of codons (ENC) (Nasrullah et al., 2015; Novembre, 2000). The ENC index normally lies between 20 and 61, where a lower value indicates extreme bias in codon usage and vice versa (Comeron and Aguade, 1998; Wright, 1990). An ENC and GC3s plot was produced to highlight the role of dominant mutation in codon usage pattern. For correlation purposes, the expected ENC values were computed for various GC3 using the method of Singh et al. (2016):

$$\text{ENC}^{\text{expected}} = 2 + s + \left(\frac{29}{s^2 + (1-s)^2} \right)$$

where 's' represents G + C contents at the 3rd codon position (GC3s).

2.4. Discerning the similarity effect of codon bias

The RSCU value of each codon was used to determine the similarity influence between the organisms in this study. All codons (except UAG, UGG, UAA, UAG and UGA) were organized in a matrix of $N \times M$ dimensions, where N is the number of species and M is the number of degenerated codons. Hierarchical clustering of this matrix was conducted based on Spearman's correlational distance of RSCU values

using Bioconductor with Ward's method (Gentleman et al., 2004). The resulting dendrogram was extracted using ggplot2 in R (Wickham, 2016). A null model was derived using a synonymous codon-shuffling algorithm for each concatenated genome sequence. Furthermore, the observed probability of Spearman's correlational distance was used as the *P*-value to represent significance.

2.5. Codon adaptation index (CAI) analysis

A quantitative method, the codon adaptation index (CAI; http://genomes.urv.es/CAIcal), was performed to determine the codon usage preferences in CCHFV, considering *H. sapiens*, *B. taurus*, *O. aries*, and *Hyalomma* as references (Puigbò et al., 2008a; Sharp and Li, 1986). This method was used to pre-determine the gene expression level based on codon sequence. The method identified the contrast in a given codon usage of CCHFV with *H. sapiens*, *B. taurus*, *O. aries*, and *Hyalomma*. CAI was used to confirm whether the CCHFV coding sequences were over fitted or less fitted to the codon usage of the reference datasets than the genes that describe the related dataset itself. Datasets of human genes were selected randomly from the Ensembl database (http://www.ensembl.org). Student's *t*-test was used to detect significant differences between CAI values obtained from various comparisons.

In order to determine if the statistically significant differences in the CAI values arose from codon preferences, we used e-CAI (http://genomes.urv.es/CAIcal/E-CA) (Puigbò et al., 2008b) to calculate the expected value of CAI (e-CAI) at the 95% confidence interval. A Kolmogorov–Smirnov test was used to calculate e-CAI (Puigbò et al., 2008b). The RSCU values of hosts *H. sapiens*, *B. taurus*, *O. aries*, and *Hyalomma* were obtained from the recently updated codon usage database (https://hive.biochemistry.gwu.edu/dna.cgi?cmd=refseq_processor&id=545408) (Athey et al., 2017).

2.6. Correspondence analysis

A multivariate statistical analysis was performed to detect variable and sample relationships. Correspondence analysis (COA) displays sets of rows and columns in a particular data set (Greenacre, 1984; Wong et al., 2010). Every ORF is denoted in 59 dimensions (59 codons) and every dimension is equal to the RSCU value of one codon (eliminating 5 codons). In order to determine the tendency within a data set, relative inertia and gene orders were measured according to their positions on distinct axes (Tao et al., 2009). The CodonW program was utilized for correspondence analysis based on the RSCU values.

2.7. Phylogenetic analysis of CCHFV

The phylogenetic tree was constructed using the maximum likelihood method in Clustal × 2 (http://www.clustal.org/clustal2/). The tree was designed using the online tool the Interactive Tree Of Life version 3 (http://itol.embl.de/) (Letunic and Bork, 2011; Serres-Giardi et al., 2012). A total of 179 ORFs strains were used in this study.

2.8. Correlation analysis

Correlation analysis was performed to describe the relationships between nucleotide contents and codon usage patterns of CCHFV. Correlation analyses were carried out by Spearman's rank correlation method (Wu et al., 2015). All statistical procedures were performed using the R corrplot package (http://rpubs.com/melike/corrplot). To examine the codon usage indices CodonW (1.4.4) software was used.

2.9. tRNA adaptation index

The tRNA adaptation index (tAI) is used to estimate tRNA usage for the coding sequences of a species (Liu et al., 2017). The tAI value of CCHFV polyprotein-coding region based on the tRNA copy number of *H. sapiens* was calculated using Visual Gene Developer (Jung and McDonald, 2011).

3. Results and discussion

3.1. Nucleotide contents analysis in CCHFV

Codon usage bias can be greatly influenced by the overall nucleotide content of the genome (Jenkins and Holmes, 2003; Nasrullah et al., 2015). A previous study suggested that nucleotide bias was the important factor of the virus-specific codon usages, thus limiting the role of codon selection and translational control (van Hemert et al., 2016). Therefore, we first determined the nucleotide compositions of the CCHFV genome to highlight the potential influence of the nucleotide constraints on codon usage. Our results indicated that the mean compositions of nucleotides A (31.34% ± 0.86) and U (24.22% ± 1.41), were significantly high in frequency, followed by C (21.48% ± 1.37) and G (22.95% ± 1.23) (Table 1, Fig. S1 (A), *t*-test, *P* < 0.01). This result was consistent with the prior studies wherein A and U frequencies were higher than C and G frequencies for avian rotaviruses and flaviviruses including dengue virus, West Nile virus, Japanese encephalitis virus, yellow fever virus, and hepatitis C virus (Kattoor et al., 2015; Lara-Ramírez et al., 2014; Moratorio et al., 2013; van Hemert and Berkhout, 2016). However, the biological causes for increased A and decreased G are unknown so it is important to determine the causes of these trends in viral RNA genomes (van Hemert et al., 2016).

To estimate the magnitude of codon bias in CCHFV, mean values for all triplets were considered during the study. The percentages of nucleotides at the third codon position were: 26.66% ± 4.04 A; 26.10% ± 1.51 U; 25.35% ± 2.74C; and 21.89% ± 2.52 G. These values were different from the expected total nucleotide contents: A (31.34%), U (24.22%), C (21.48%), and G (22.95%); in particular, the percentage of A3 was lower than the expected percentage of A, while the percentage of C3 was higher than the expected percentage of C, suggesting that A3, U3, C3, and G3 may influence selective pressure. In order to confirm our hypothesis, we determined the tRNA adaptation index (tAI) (the mean tAI values: A3 (0.237), U3 (0.370), G3 (0.360) and C3 (0.430)), for the reason that the tAI values indicate the natural selection on the third position's nucleotide contents. The higher the tAI value, the higher the third position's nucleotide contents would be, and vice versa. Thus, the fact that the percentage of A3 was lower than the expected percentage of A, and the percentage of C3 was higher than the expected percentage of C, is explained by the observation that the relative low tAI values of A3 (0.237), and the high tAI value of C3 (0.430). Therefore, we suggest an influence of selection pressure on shaping the codon usage pattern of CCHFV (Table 1, Fig. S1 (B)).

GC content at each codon position is assumed to be a good indicator of base composition bias. GC nucleotide content ranges were as follows: 43.21% to 50.31% (mean = 46.56, SD = 2.42) at the 1st codon position; 35.12% to 45% (mean = 39.5, SD = 3.61) at the 2nd codon position; and 40.07% to 45.12% (mean = 43.03, SD = 1.66) at the 1st and 2nd codon positions. The mean AU and GC contents were 55.57% ± 2.19 and 44.43% ± 2.19, respectively, and the mean AU3 and GC3 contents were 52.76% ± 4.44 and 47.24% ± 4.44, respectively, highlighting that A and U nucleotides are more likely to occur at the end of codons (Table 1, Fig. S1 (C, D)). Taken together, these data showed that a substantial portion of CCHFV genomes are composed of A and U nucleotides, which is consistent with previous reports (Nasrullah et al., 2015). Previous studies have also suggested that the nucleotide composition of first and second codon positions in a gene depends on the amino acid composition of the protein product, and its variation is constrained by functional selection at the level of protein evolution. At the third codon positions of a viral gene, 69% of the possible alterations represent synonymous or silent mutations, which are not restricted by functional selection of amino acids (van Hemert et al., 2016).

Table 1
Nucleotide composition analysis of CCHFV coding sequences (%).

Sequences/ parameters	A	C	U	G	GC	AU	GC1	GC2	A3	C3	U3	G3	ENC	GC12	GC3	AU3	ARO	Gravy
JX908640.1	30.64	21.26	23.88	24.22	45.48	54.52	48.86	37.68	21.53	25.67	28.57	24.22	54.01	43.27	49.90	50.10	0.10	−0.40
HM452305.1	30.23	22.15	22.98	24.64	46.79	53.21	49.28	37.89	20.70	28.16	26.09	25.05	52.69	43.59	53.21	46.79	0.10	−0.40
GU477489.1	31.19	22.15	22.91	23.74	45.89	54.11	49.48	37.27	22.98	27.74	26.09	23.19	54.43	43.38	50.93	49.07	0.10	−0.41
AY277676.2	31.26	22.08	22.98	23.67	45.76	54.24	49.48	37.27	23.19	27.54	26.29	22.98	55.02	43.38	50.52	49.48	0.10	−0.41
FJ562093.1	30.57	22.29	22.50	24.64	46.93	53.07	49.69	37.89	21.74	28.16	25.05	25.05	53.90	43.79	53.21	46.79	0.10	−0.42
GU477494.1	30.43	22.02	22.84	24.71	46.72	53.28	49.28	38.30	21.74	27.74	25.67	24.84	54.01	43.79	52.59	47.41	0.10	−0.41
KY484031.1	31.19	22.71	21.74	24.36	47.07	52.93	50.10	37.47	23.60	29.61	22.77	24.02	55.27	43.79	53.62	46.38	0.10	−0.40
KY354080.1	30.64	22.15	22.64	24.57	46.72	53.28	49.48	38.10	22.15	27.95	25.26	24.64	54.91	43.69	52.59	47.41	0.10	−0.42
AF362080.1	31.06	22.64	22.08	24.22	46.86	53.14	49.48	37.89	22.98	29.19	23.81	24.02	53.66	43.69	53.21	46.79	0.10	−0.42
DQ227496.1	30.43	22.50	22.29	24.78	47.27	52.73	49.90	37.89	21.33	28.57	24.64	25.47	53.88	43.90	54.04	45.96	0.10	−0.42
DQ227495.1	30.43	22.57	22.22	24.78	47.34	52.66	49.90	38.10	21.33	28.57	24.64	25.47	54.06	44.00	54.04	45.96	0.10	−0.43
DQ217602.1	30.50	22.50	22.29	24.71	47.20	52.80	49.90	37.89	21.53	28.57	24.64	25.26	53.91	43.90	53.83	46.17	0.10	−0.42
DQ211642.1	31.06	22.50	22.15	24.29	46.79	53.21	49.69	37.47	22.98	29.19	23.81	24.02	53.97	43.58	53.21	46.79	0.10	−0.41
KY484048.1	29.40	22.08	23.19	25.33	47.41	52.59	49.90	38.51	19.25	27.74	26.92	26.09	53.51	44.21	53.83	46.17	0.10	−0.41
DQ144418.1	29.40	22.08	23.19	25.33	47.41	52.59	49.90	38.51	19.25	27.74	26.92	26.09	53.51	44.21	53.83	46.17	0.10	−0.41
DQ211650.1	29.40	22.08	23.19	25.33	47.41	52.59	49.90	38.51	19.25	27.74	26.92	26.09	53.51	44.21	53.83	46.17	0.10	−0.41
EU871766.2	30.99	22.43	22.71	23.88	46.31	53.69	49.48	37.68	22.57	28.36	25.67	23.40	54.33	43.58	51.76	48.24	0.10	−0.42
DQ211638.1	30.16	22.98	21.88	24.98	47.96	52.04	48.24	37.89	18.63	30.02	23.60	27.74	53.20	43.07	57.76	42.24	0.10	−0.45
KY213714.1	30.23	21.95	23.19	24.64	46.58	53.42	49.28	38.10	20.91	27.54	26.71	24.84	52.75	43.69	52.38	47.62	0.10	−0.40
JN572089.1	30.85	21.33	23.33	24.50	45.82	54.18	49.48	37.89	22.36	25.47	27.54	24.64	54.82	43.69	50.10	49.90	0.10	−0.42
JN572088.1	30.78	21.60	23.05	24.57	46.17	53.83	49.69	37.89	22.36	26.29	26.71	24.64	55.16	43.79	50.93	49.07	0.10	−0.42
JN572087.1	30.78	21.60	23.05	24.57	46.17	53.83	49.69	37.89	22.36	26.29	26.71	24.64	55.16	43.79	50.93	49.07	0.10	−0.42
JX051650.1	30.78	21.60	23.05	24.57	46.17	53.83	49.69	37.89	22.36	26.29	26.71	24.64	54.74	43.79	50.93	49.07	0.10	−0.42
KT384401.1	30.16	21.95	23.19	24.71	46.65	53.35	49.28	38.30	20.91	27.54	26.71	24.84	52.69	43.79	52.38	47.62	0.10	−0.40
JF922681.1	30.85	21.46	23.19	24.50	45.96	54.04	49.69	37.89	22.57	25.88	27.12	24.43	54.98	43.79	50.31	49.69	0.10	−0.42
JF922679.1	30.71	21.53	23.12	24.64	46.17	53.83	49.69	37.89	22.15	26.09	26.92	24.84	54.84	43.79	50.93	49.07	0.10	−0.42
JF922674.1	30.78	21.60	23.05	24.57	46.17	53.83	49.69	37.89	22.36	26.29	26.71	24.64	55.16	43.79	50.93	49.07	0.10	−0.42
KC867274.1	30.23	21.95	23.19	24.64	46.58	53.42	49.28	38.10	20.91	27.54	26.71	24.84	52.75	43.69	52.38	47.62	0.10	−0.40
KJ676542.1	31.19	21.67	23.05	24.09	45.76	54.24	49.69	37.89	23.60	26.50	26.71	23.19	54.99	43.79	49.69	50.31	0.10	−0.42
KJ566219.1	30.23	22.22	22.98	24.57	46.79	53.21	50.21	37.97	21.78	28.42	25.31	24.48	52.16	43.59	53.21	46.79	0.10	−0.39
KJ196326.1	30.30	22.08	23.05	24.57	46.65	53.35	49.28	37.89	20.91	27.95	26.29	24.84	52.72	43.59	52.80	47.20	0.10	−0.40
KJ027522.1	30.57	22.57	22.43	24.43	47.00	53.00	49.69	37.68	20.91	28.78	26.29	24.43	54.05	43.69	52.80	46.38	0.10	−0.42
KJ027521.1	30.57	22.36	22.71	24.36	46.72	53.28	49.28	38.10	21.74	28.36	25.47	24.84	54.04	43.69	52.80	47.20	0.10	−0.41
DQ446214.1	30.23	22.08	23.12	24.57	46.65	53.35	49.28	37.89	21.12	27.95	26.50	24.84	52.64	43.59	52.80	47.20	0.10	−0.40
DQ446213.1	30.23	22.02	23.19	24.57	46.58	53.42	49.28	37.89	20.70	27.74	26.71	24.84	52.63	43.59	52.59	47.41	0.10	−0.40
DQ446212.1	30.30	22.08	23.05	24.57	46.65	53.35	49.28	38.10	21.12	27.95	26.29	24.64	52.56	43.69	52.59	47.41	0.10	−0.40
DQ211641.1	30.71	22.64	22.36	24.29	46.93	53.07	50.31	37.89	21.95	28.57	25.47	24.02	54.40	44.10	52.59	47.41	0.10	−0.43
KJ682820.1	29.81	22.50	22.29	25.40	47.90	52.10	50.31	38.10	19.67	28.57	25.05	26.71	52.76	44.21	55.28	44.72	0.10	−0.43
KY484036.1	30.50	22.43	22.57	24.50	46.93	53.07	49.90	38.10	21.33	28.36	25.88	24.43	53.30	44.00	52.80	47.20	0.10	−0.41
KX013452.1	30.78	20.29	24.09	24.84	45.13	54.87	50.31	37.68	22.77	22.15	29.81	25.26	53.24	44.00	47.41	52.59	0.09	−0.40
KX238958.1	30.71	22.41	22.27	24.62	47.03	52.97	50.21	37.97	21.78	28.42	25.31	24.48	53.47	44.09	52.90	47.10	0.10	−0.43
KY362516.1	30.50	22.15	22.98	24.36	46.51	53.49	49.07	37.68	21.12	28.16	26.09	24.64	53.38	43.38	52.80	47.20	0.10	−0.40
DQ211645.1	30.43	22.22	23.05	24.36	46.51	53.49	49.07	37.68	20.91	28.16	26.29	24.64	53.34	43.38	52.80	47.20	0.10	−0.40
KY484037.1	30.23	22.71	22.98	24.57	46.79	53.21	49.28	38.10	20.70	28.36	25.88	25.05	51.90	43.48	53.42	46.58	0.10	−0.39
DQ206447.1	30.50	22.08	22.98	24.43	47.14	52.86	49.48	37.68	21.12	29.19	24.64	25.05	54.24	43.58	54.24	45.76	0.10	−0.42
KY484028.1	30.71	22.71	22.36	24.09	46.45	53.55	49.28	37.47	21.33	28.16	26.09	24.43	54.86	43.38	54.24	47.41	0.10	−0.43
KU161587.1	30.85	22.71	22.36	24.09	46.79	53.21	49.48	37.68	22.15	29.19	24.64	24.02	53.50	43.58	53.21	46.79	0.10	−0.42
KU161586.1	31.12	22.57	22.29	23.88	46.58	53.42	49.48	37.97	22.77	28.99	24.84	23.40	53.73	43.69	52.38	47.62	0.10	−0.44
KR814855.1	30.50	21.95	23.05	24.43	47.00	53.00	49.48	37.68	21.12	28.78	25.05	25.05	54.14	43.58	53.83	46.17	0.10	−0.42
KR814854.1	30.85	22.57	22.50	23.95	46.10	53.90	49.48	37.68	22.57	26.92	26.71	24.22	54.61	43.58	51.14	48.86	0.10	−0.42
KR814853.1	30.99	22.57	22.36	23.60	46.51	53.49	49.48	37.68	22.15	28.78	25.05	23.60	55.12	43.58	52.38	47.62	0.10	−0.42
KR814852.1	30.57	22.71	22.36	24.36	47.07	52.93	49.48	37.68	21.33	29.19	24.64	24.84	54.61	43.58	54.04	45.96	0.10	−0.42
KR814851.1	30.64	22.57	22.50	24.29	46.86	53.14	49.28	37.68	21.33	28.78	25.05	24.84	53.61	43.48	53.62	46.38	0.10	−0.42

Table 1 (continued)

Sequences/parameters	A	C	U	G	GC	AU	GC1	GC2	A3	C3	U3	G3	ENC	GC12	GC3	AU3	ARO	Gravy
KR814850.1	30.57	22.50	22.57	24.36	46.86	53.14	49.48	37.68	21.33	28.57	25.26	24.84	53.62	43.58	53.42	46.58	0.10	−0.42
KY484027.1	31.06	23.05	21.81	24.09	47.14	52.86	50.31	37.47	22.36	30.02	24.02	23.60	50.17	43.89	53.62	46.38	0.10	−0.45
DQ211640.1	30.92	23.19	21.81	24.09	47.27	52.73	50.31	37.27	21.74	30.43	24.02	23.81	49.99	43.79	54.24	45.76	0.10	−0.45
DQ211639.1	31.06	23.05	21.81	24.09	47.14	52.86	50.31	37.47	22.36	30.02	24.02	23.60	50.17	43.89	53.62	46.38	0.10	−0.45
KY362517.1	30.99	22.43	22.64	23.95	46.38	53.62	49.48	37.68	22.57	28.36	25.47	23.60	53.83	43.58	51.97	48.03	0.10	−0.42
GQ337053.1	30.71	22.15	22.77	24.36	46.51	53.49	49.69	37.89	22.15	27.74	25.88	24.22	53.96	43.79	51.97	48.03	0.10	−0.43
DQ211649.1	30.71	22.57	22.43	24.29	46.86	53.14	49.69	37.68	21.95	28.78	24.84	24.43	54.01	43.69	53.21	46.79	0.10	−0.42
KY484044.1	30.23	22.29	22.29	25.19	47.48	52.52	50.10	38.10	20.50	27.95	25.26	26.29	52.91	44.10	54.24	45.76	0.10	−0.43
DQ076413.1	30.16	21.67	23.60	24.57	46.24	53.76	49.69	38.72	21.53	26.50	28.16	23.81	54.24	44.21	50.31	49.69	0.10	−0.40
DQ076415.1	30.23	22.29	22.36	25.12	47.41	52.59	50.10	37.89	20.50	27.95	25.26	26.29	53.14	44.00	54.24	45.76	0.10	−0.43
KX013485.1	29.95	21.81	23.53	24.71	46.51	53.49	49.90	38.72	20.91	26.71	28.16	24.22	54.08	44.31	50.93	49.07	0.10	−0.40
KX013449.1	30.57	21.88	22.77	24.78	46.65	53.35	49.28	37.89	21.74	27.33	25.47	25.47	55.21	43.59	52.80	47.20	0.10	−0.43
AF481799.1	30.78	22.57	22.02	24.64	47.20	52.80	49.69	37.68	22.36	29.40	23.40	24.84	55.24	43.69	54.24	45.76	0.10	−0.42
DQ133507.1	30.64	22.43	22.57	24.36	46.79	53.21	49.69	37.68	21.74	28.36	25.26	24.64	54.00	43.69	53.00	47.00	0.10	−0.42
AY223476.1	31.01	22.09	24.79	22.11	44.20	55.80	44.44	43.85	30.06	25.68	25.62	18.64	51.20	44.15	44.32	55.68	0.07	−0.16
AY900141.1	31.37	21.69	24.96	21.98	43.67	56.33	44.50	43.97	31.08	24.66	26.38	17.88	51.02	44.24	42.54	57.46	0.07	−0.18
AY900142.1	30.86	21.92	24.65	22.57	44.49	55.51	44.93	44.45	29.55	24.63	26.35	19.47	50.26	44.69	44.09	55.91	0.07	−0.19
AY900143.1	30.76	22.73	23.84	22.67	45.39	54.61	46.53	43.71	29.35	25.76	24.71	20.18	52.46	45.12	45.94	54.06	0.07	−0.12
AY900144.1	30.61	22.34	24.63	22.42	44.76	55.24	45.00	44.46	29.01	25.75	26.17	19.06	50.76	44.73	44.82	55.18	0.07	−0.20
AY900145.1	31.28	21.62	25.29	21.81	43.43	56.57	43.37	43.43	30.53	24.97	25.98	18.52	50.42	43.40	43.49	56.51	0.07	−0.14
DQ019222.1	30.78	22.71	23.86	22.65	45.35	54.65	46.47	43.71	29.41	25.76	24.71	20.12	52.44	45.09	45.88	54.12	0.07	−0.12
DQ094832.1	30.55	22.73	24.02	22.71	45.43	54.57	45.82	43.47	28.29	25.88	24.71	21.12	52.56	44.65	47.00	53.00	0.07	−0.14
DQ206448.1	31.26	22.58	24.10	22.06	44.64	55.36	44.17	44.58	30.14	25.52	24.69	19.66	50.87	44.38	45.17	54.83	0.07	−0.19
DQ211625.1	30.29	23.17	23.82	22.72	45.89	54.11	45.22	43.81	27.18	27.30	24.17	21.34	51.73	44.52	48.64	51.36	0.07	−0.18
DQ211626.1	31.92	21.58	24.77	21.73	43.31	56.69	43.21	43.56	31.04	25.75	25.81	18.87	51.44	43.39	43.15	56.85	0.08	−0.16
DQ211627.1	31.86	21.75	24.61	21.77	43.52	56.48	43.39	43.86	30.98	24.28	25.69	18.81	52.17	43.63	43.33	56.67	0.07	−0.16
DQ211628.1	31.15	23.11	23.84	21.89	45.01	54.99	44.22	44.46	28.95	24.51	24.71	19.34	51.13	44.34	46.34	53.66	0.07	−0.20
DQ211629.1	31.22	21.62	25.29	21.87	43.49	56.51	43.49	43.31	30.36	27.00	25.98	18.70	50.40	43.40	43.67	56.33	0.07	−0.15
DQ211630.1	31.24	22.48	24.12	22.16	44.64	55.36	44.46	43.99	30.14	24.97	24.39	19.60	51.39	44.23	45.47	54.53	0.07	−0.17
DQ211631.1	31.28	22.64	24.00	22.08	44.72	55.28	44.23	44.70	30.20	25.87	24.57	19.66	50.83	44.47	45.23	54.77	0.07	−0.19
DQ211632.1	31.71	21.88	24.78	21.62	43.50	56.50	45.14	43.60	32.11	25.58	26.13	17.12	49.95	44.37	41.77	58.23	0.07	−0.16
DQ211633.1	30.95	22.72	24.08	22.25	44.97	55.03	44.08	44.91	30.18	24.64	23.91	18.88	50.86	44.50	45.92	54.08	0.07	−0.19
DQ211634.1	30.80	21.78	24.87	22.55	44.33	55.67	45.46	44.57	30.09	27.04	26.94	18.87	49.95	45.02	42.97	57.03	0.07	−0.18
DQ211635.1	31.12	22.41	24.36	22.11	44.52	55.48	44.08	44.32	30.24	24.09	24.62	18.99	50.70	44.20	45.15	54.85	0.07	−0.18
DQ211636.1	31.26	22.52	24.10	22.12	44.64	55.36	44.35	44.76	30.43	26.15	24.75	19.36	51.30	44.56	44.82	55.18	0.07	−0.19
DQ211637.1	30.76	22.71	23.86	22.67	45.37	54.63	46.47	43.71	29.35	25.46	24.71	20.18	52.48	45.09	45.94	54.06	0.07	−0.12
DQ446215.1	31.12	21.54	25.25	22.02	43.56	56.44	44.51	43.62	30.09	25.76	27.36	18.46	51.58	44.07	42.55	57.45	0.08	−0.18
DQ446216.1	31.08	21.58	25.34	22.00	43.58	56.42	44.57	43.80	30.21	24.09	27.42	18.22	51.47	44.19	42.37	57.63	0.08	−0.17
EU037902.1	31.10	22.77	23.84	22.28	45.06	54.94	43.93	44.52	29.31	26.29	23.98	20.43	51.60	44.23	46.71	53.29	0.07	−0.21
FJ562094.1	31.39	21.99	24.48	22.14	44.13	55.87	44.82	44.05	30.51	24.20	25.97	19.32	51.89	44.44	43.52	56.48	0.07	−0.17
GQ337054.1	30.97	22.77	24.02	22.24	45.02	54.98	44.58	45.00	30.02	25.75	24.51	19.72	51.40	44.79	45.47	54.53	0.07	−0.18
GU477490.1	31.50	22.40	24.08	22.02	44.42	55.58	44.23	44.70	30.73	25.16	24.93	19.18	51.33	44.47	44.35	55.65	0.07	−0.20
GU477493.1	31.31	21.99	24.46	22.24	44.23	55.77	44.94	43.93	30.45	24.44	25.74	19.38	52.06	44.44	43.82	56.18	0.07	−0.16
HM452306.1	31.16	21.46	25.42	21.96	43.42	56.58	44.45	43.68	30.15	23.86	27.72	18.28	51.75	44.07	42.14	57.86	0.08	−0.18
JN572083.1	31.30	21.36	25.44	21.90	43.26	56.74	44.81	43.68	30.68	23.20	28.01	18.10	51.81	44.25	41.31	58.69	0.07	−0.19
JN572084.1	31.30	21.40	25.42	21.88	43.28	56.72	44.75	43.68	30.68	23.32	27.89	18.10	51.84	44.22	41.42	58.58	0.07	−0.19
JN572085.1	31.32	21.35	25.48	21.86	43.20	56.80	44.75	43.62	30.68	23.20	28.07	18.04	51.77	44.19	41.25	58.75	0.07	−0.18
JN572086.1	31.30	21.40	25.40	21.90	43.30	56.70	44.81	43.68	30.68	23.32	27.89	18.10	51.84	44.25	41.42	58.58	0.07	−0.19
KC867273.1	31.16	21.52	25.46	21.86	43.38	56.62	44.39	43.68	30.33	24.09	27.60	17.98	51.70	44.04	42.08	57.92	0.08	−0.18
KJ682805.1	31.83	22.39	24.38	21.40	43.79	56.21	43.73	43.08	30.95	26.39	24.50	18.17	50.83	43.41	44.56	55.44	0.07	−0.18
KJ682806.1	31.62	22.21	24.62	21.56	43.77	56.23	43.96	42.96	30.71	25.98	24.91	18.40	50.91	43.46	44.38	55.62	0.07	−0.17
KJ682807.1	31.68	22.41	24.34	21.58	43.98	56.02	44.02	43.08	30.65	26.39	24.50	18.46	50.92	43.55	44.85	55.15	0.07	−0.19
KJ682808.1	31.76	22.33	24.54	21.38	43.71	56.29	43.67	42.78	31.07	26.63	24.26	18.05	51.22	43.23	44.67	55.33	0.07	−0.17
KJ682809.1	31.40	22.14	24.63	21.83	43.97	56.03	43.81	43.40	30.20	25.75	25.10	18.95	50.92	43.61	44.70	55.30	0.07	−0.18

Table 1 (continued)

Sequences/parameters	A	C	U	G	GC	AU	GC1	GC2	A3	C3	U3	G3	ENC	GC12	GC3	AU3	ARO	Gravy
KJ682810.1	31.87	22.01	24.60	21.52	43.53	56.47	43.85	42.96	31.30	25.56	24.91	18.22	50.68	43.41	43.79	56.21	0.07	−0.18
KJ682811.1	30.95	22.74	24.06	22.25	44.99	55.01	44.08	44.97	30.18	27.04	23.91	18.88	50.90	44.53	45.92	54.08	0.07	−0.19
KJ682812.1	31.12	22.47	24.28	22.13	44.60	55.40	44.14	44.50	30.24	26.15	24.62	18.99	50.74	44.32	45.15	54.85	0.07	−0.19
KJ682813.1	31.22	22.00	24.85	21.94	43.94	56.06	44.21	43.92	30.33	25.52	25.99	18.16	51.53	44.07	43.68	56.32	0.07	−0.18
KJ682814.1	31.34	21.54	25.28	21.84	43.38	56.62	44.27	43.74	30.80	24.27	27.06	17.86	50.84	44.01	42.14	57.86	0.07	−0.17
KR814856.1	31.32	22.72	23.92	22.04	44.76	55.24	44.35	44.70	30.20	25.64	24.57	19.60	50.96	44.53	45.23	54.77	0.07	−0.20
KR814857.1	31.32	22.56	24.10	22.02	44.58	55.42	44.23	44.29	30.31	25.87	24.45	19.36	51.63	44.26	45.23	54.77	0.07	−0.19
KR814858.1	31.32	22.74	23.92	22.02	44.76	55.24	44.40	44.64	30.20	25.70	24.57	19.54	50.87	44.52	45.23	54.77	0.07	−0.20
KR814859.1	31.34	22.58	24.08	22.01	44.58	55.42	44.29	44.29	30.37	25.87	24.45	19.30	51.65	44.29	45.17	54.83	0.07	−0.19
KR814860.1	31.22	22.85	23.82	22.10	44.96	55.04	44.58	44.52	30.20	26.17	24.04	19.60	51.00	44.55	45.77	54.23	0.07	−0.18
KR814861.1	31.36	22.66	23.96	22.02	44.68	55.32	44.35	44.52	30.43	25.87	24.39	19.30	51.62	44.44	45.17	54.83	0.07	−0.19
KR814862.1	31.30	22.76	23.92	22.02	44.78	55.22	44.58	44.70	30.31	25.64	24.63	19.42	50.85	44.64	45.06	54.94	0.07	−0.20
KR814863.1	31.32	22.52	24.08	22.08	44.60	55.40	44.17	44.40	30.20	25.52	24.57	19.72	50.67	44.29	45.23	54.77	0.07	−0.19
KR814864.1	31.36	22.85	23.78	22.01	44.86	55.14	44.23	44.70	30.43	26.11	23.92	19.54	50.92	44.47	45.65	54.35	0.07	−0.20
KR814865.1	31.26	22.89	23.76	22.08	44.98	55.02	44.40	44.52	30.14	26.29	23.86	19.72	50.89	44.46	46.00	54.00	0.07	−0.19
KR814866.1	31.34	22.60	24.02	22.04	44.64	55.36	44.17	44.23	30.25	26.11	24.22	19.42	51.60	44.20	45.53	54.47	0.07	−0.18
KR814867.1	31.18	22.64	24.02	22.16	44.80	55.20	44.35	44.58	30.08	25.70	24.45	19.78	50.82	44.47	45.47	54.53	0.07	−0.20
KR814868.1	31.22	22.67	23.97	22.14	44.81	55.19	44.31	44.25	30.04	26.30	24.11	19.55	51.72	44.28	45.85	54.15	0.07	−0.18
KR814869.1	31.34	22.85	23.80	22.01	44.86	55.14	44.17	44.82	30.31	25.99	24.10	19.60	50.94	44.50	45.59	54.41	0.07	−0.21
KR814870.1	31.36	22.81	23.80	22.02	44.84	55.16	44.11	44.70	30.37	26.05	23.92	19.66	50.93	44.41	45.71	54.29	0.07	−0.20
KR814871.1	31.24	22.53	24.05	22.18	44.71	55.29	44.14	44.49	29.98	25.83	24.53	19.67	51.77	44.32	45.50	54.50	0.07	−0.19
KR814872.1	31.24	22.56	24.04	22.16	44.72	55.28	44.52	44.40	30.31	25.81	24.45	19.42	51.77	44.46	45.23	54.77	0.07	−0.18
KR814873.1	31.32	22.58	24.02	22.08	44.66	55.34	44.29	44.05	30.20	26.11	24.16	19.54	51.76	44.17	45.65	54.35	0.07	−0.18
KR814874.1	31.36	22.77	23.88	21.99	44.76	55.24	44.29	44.70	30.31	25.81	24.39	19.48	51.01	44.50	45.29	54.71	0.07	−0.20
NC_005300.2	31.55	22.02	24.65	21.78	43.80	56.20	44.57	44.04	31.39	25.16	25.82	17.63	50.59	44.31	42.79	57.21	0.07	−0.18
KC344855.1	32.30	19.25	26.25	22.20	41.45	58.55	45.67	35.38	28.16	21.06	28.54	22.25	51.82	40.53	43.31	56.69	0.08	−0.27
HM452307.1	32.51	19.11	26.37	22.01	41.12	58.88	45.34	35.43	28.94	21.06	28.46	21.54	51.68	40.39	42.60	57.40	0.08	−0.28
GU477492.1	32.48	19.49	25.94	22.09	41.58	58.42	45.62	35.25	28.46	21.64	27.67	22.23	52.22	40.44	43.87	56.13	0.08	−0.27
KY484033.1	32.74	18.99	26.42	21.84	40.83	59.17	45.18	35.50	29.55	20.60	28.64	21.21	51.13	40.34	41.81	58.19	0.07	−0.28
KY354082.1	32.51	19.54	25.92	22.03	41.57	58.43	45.62	35.12	28.54	21.82	27.50	22.15	52.63	40.37	43.97	56.03	0.08	−0.27
FJ562095.1	32.42	19.50	25.96	22.12	41.62	58.38	45.67	35.20	28.38	21.74	27.62	22.25	52.53	40.44	43.99	56.01	0.08	−0.27
DQ211616.1	32.75	18.99	26.42	21.84	40.83	59.17	45.18	35.48	29.55	20.60	28.64	21.21	51.14	40.33	41.81	58.19	0.07	−0.28
KY484046.1	32.31	19.61	25.98	22.10	41.71	58.29	45.39	35.45	27.62	21.77	28.08	22.53	51.40	40.42	44.30	55.70	0.08	−0.29
DQ099335.2	32.32	19.63	25.96	22.09	41.72	58.28	45.41	35.43	27.62	21.79	28.05	22.53	51.41	40.42	44.32	55.68	0.08	−0.29
DQ211624.1	32.34	19.61	25.98	22.07	41.69	58.31	45.36	35.43	27.65	21.77	28.08	22.50	51.37	40.40	44.27	55.73	0.08	−0.29
DQ211612.1	32.42	20.16	24.88	22.54	42.70	57.30	46.30	35.81	28.51	23.01	25.49	22.99	51.37	41.06	46.00	54.00	0.07	−0.28
KY213712.1	32.30	19.29	26.27	22.14	41.43	58.57	45.65	35.39	28.11	21.14	28.64	22.10	51.74	40.52	43.24	56.76	0.08	−0.27
JN572092.1	32.60	19.11	26.35	21.95	41.05	58.95	45.21	35.40	29.04	20.96	28.41	21.59	51.51	40.31	42.55	57.45	0.08	−0.28
JN572091.1	32.61	19.07	26.37	21.95	41.02	58.98	45.29	35.38	29.12	20.91	28.48	21.49	51.53	40.34	42.40	57.60	0.08	−0.28
JN572090.1	32.59	19.09	26.34	21.98	41.07	58.93	45.18	35.38	28.94	20.98	28.41	21.67	51.58	40.28	42.65	57.35	0.08	−0.28
JN627865.1	32.57	19.12	26.36	21.95	41.07	58.93	45.21	35.40	28.97	20.98	28.43	21.62	51.56	40.31	42.60	57.40	0.08	−0.28
KC867272.1	32.54	19.03	26.41	22.02	41.05	58.95	45.44	35.40	28.97	20.68	28.71	21.64	51.70	40.42	42.32	57.68	0.08	−0.28
KX238956.1	32.77	19.09	26.12	21.88	41.11	58.89	44.98	35.55	29.65	21.44	27.55	21.36	52.15	40.27	42.80	57.20	0.08	−0.28
KY362514.1	32.73	19.13	26.03	21.99	41.25	58.75	44.70	35.63	29.42	21.57	27.17	21.84	52.18	40.17	43.41	56.59	0.08	−0.28
DQ211615.1	32.77	19.23	26.09	21.90	41.14	58.86	44.86	35.68	29.73	21.52	27.39	21.36	52.06	40.27	42.88	57.12	0.08	−0.28
KJ682804.1	32.73	19.26	26.09	21.90	41.14	58.86	44.86	35.68	29.73	21.52	27.39	21.36	52.06	40.27	42.88	57.12	0.08	−0.28
KY484034.1	32.77	19.23	26.09	21.90	41.50	58.50	45.39	35.53	28.71	21.34	27.70	22.25	51.87	40.46	43.59	56.41	0.08	−0.28
AY947891.1	32.77	19.23	26.09	21.90	40.89	59.11	44.42	35.72	29.55	21.13	27.92	21.41	51.95	40.07	42.54	57.46	0.08	−0.28
KX013450.1	32.41	19.29	26.33	22.22	41.02	58.98	45.11	35.35	29.24	21.11	28.16	21.49	51.49	40.23	42.60	57.40	0.08	−0.28
DQ211619.1	32.67	19.09	26.38	21.95	41.14	59.03	45.11	35.43	29.17	21.26	27.98	21.59	51.64	40.28	42.85	57.15	0.08	−0.28
KY484039.1	32.65	19.04	26.09	21.93	40.97	58.79	45.21	35.43	29.24	20.86	28.48	21.41	51.57	40.32	42.27	57.73	0.08	−0.28
AY675240.2	32.70	19.28	26.20	21.93	41.21	58.73	44.88	35.61	29.27	21.57	27.60	21.57	51.83	40.25	43.13	56.87	0.08	−0.28
KJ484030.1	32.54	19.19	26.20	22.07	41.27	58.73	44.73	35.53	28.79	21.46	27.67	22.07	51.88	40.13	43.54	56.46	0.08	−0.27
KU161583.1	32.63	19.33	26.07	21.97	41.30	58.70	44.68	35.58	28.92	21.87	27.45	21.77	51.86	40.13	43.64	56.36	0.08	−0.27

(continued on next page)

Table 1 (*continued*)

Sequences/parameters	A	C	U	G	GC	AU	GC1	GC2	A3	C3	U3	G3	ENC	GC12	GC3	AU3	ARO	Gravy
KU161582.1	32.62	19.34	26.03	22.01	41.36	58.64	44.68	35.55	28.89	22.00	27.27	21.84	52.04	40.12	43.84	56.16	0.08	− 0.27
AY995166.2	32.62	19.24	26.14	22.00	41.24	58.76	44.78	35.50	28.94	21.59	27.62	21.84	51.98	40.14	43.44	56.56	0.08	− 0.27
KR814893.1	32.62	19.23	26.14	22.01	41.24	58.76	44.86	35.43	28.92	21.54	27.65	21.90	51.92	40.15	43.44	56.56	0.08	− 0.27
KR814892.1	32.63	19.23	26.16	21.98	41.21	58.79	44.68	35.61	28.99	21.57	27.67	21.77	51.99	40.15	43.34	56.66	0.08	− 0.27
KR814891.1	32.52	19.28	26.14	22.06	41.34	58.66	44.78	35.53	28.71	21.67	27.57	22.05	52.20	40.16	43.72	56.28	0.08	− 0.27
KR814890.1	32.67	19.27	26.11	21.95	41.21	58.79	44.83	35.45	29.07	21.64	27.57	21.72	51.97	40.14	43.36	56.64	0.08	− 0.28
KR814889.1	32.60	19.15	26.23	22.02	41.17	58.83	44.86	35.53	28.97	21.29	27.90	21.84	51.93	40.20	43.13	56.87	0.08	− 0.27
KY484025.1	32.85	19.24	26.07	21.84	41.08	58.92	44.80	35.61	29.83	21.59	27.34	21.24	51.89	40.21	42.83	57.17	0.08	− 0.29
DQ211614.1	32.78	19.27	26.05	21.90	41.17	58.83	44.70	35.76	29.73	21.62	27.22	21.44	52.01	40.23	43.06	56.94	0.08	− 0.28
DQ211613.1	32.85	19.24	26.07	21.84	41.08	58.92	44.80	35.61	29.83	21.59	27.34	21.24	51.89	40.21	42.83	57.17	0.08	− 0.29
KY362515.1	32.80	19.35	26.00	21.85	41.19	58.81	45.19	35.59	29.66	21.58	27.53	21.22	51.95	40.39	42.80	57.20	0.07	− 0.27
GQ337055.1	32.70	19.41	26.03	21.86	41.27	58.73	45.11	35.55	29.19	21.67	27.65	21.49	52.26	40.33	43.16	56.84	0.08	− 0.28
DQ211623.1	32.79	19.38	25.98	21.84	41.22	58.78	45.11	35.50	29.30	21.52	27.65	21.54	52.17	40.31	43.06	56.94	0.08	− 0.28
KY484043.1	32.88	19.33	25.92	21.88	41.21	58.79	45.03	35.66	29.80	21.49	27.27	21.44	51.99	40.35	42.93	57.07	0.08	− 0.29
KX013483.1	32.11	19.99	25.63	22.28	42.26	57.74	45.67	35.38	27.22	22.81	27.04	22.93	51.93	40.53	45.74	54.26	0.08	− 0.29
DQ076412.1	32.16	19.97	25.65	22.22	42.19	57.81	45.57	35.43	27.27	22.68	27.17	22.88	51.91	40.50	45.57	54.43	0.08	− 0.29
KX013447.1	32.69	19.26	26.20	21.85	41.11	58.89	45.44	35.50	29.37	21.11	28.23	21.29	52.04	40.47	42.40	57.60	0.08	− 0.27
EU044832.1	32.53	19.33	26.12	22.02	41.35	58.65	44.98	35.58	28.84	21.74	27.67	21.74	52.01	40.28	43.49	56.51	0.08	− 0.27
Mean ± STD	31.34 ± 0.86	21.48 ± 1.37	24.22 ± 1.41	22.95 ± 1.23	44.43 ± 2.19	55.57 ± 2.19	46.56 ± 2.42	39.50 ± 3.61	26.66 ± 4.04	25.35 ± 2.74	26.10 ± 1.51	21.89 ± 2.52	52.34 ± 1.36	43.03 ± 1.66	47.24 ± 4.44	52.76 ± 4.44	0.08 ± 0.01	0.00 ± 0.1

ENC represents the effective number of codons.

GC12 represents the G + C content at the first and second positions of codons.

GC3 represents the G + C content at the third positions of codons.

AU3 represents the A + U content at the third positions of codons.

Gravy represents the hydrophobicity of protein.

ARO represents the aromaticity of protein.

3.2. Determining codon usage preferences using RSCU analysis

An RSCU analysis was performed to determine the synonymous codon usage pattern in CCHFV coding sequences. The RSCU values for each synonymous codon were calculated and matched with different potential hosts such as *H. sapiens*, *Hyalomma*, *B. taurus*, and *O. aries* (Table 2, Fig. S2). Out of the eighteen most abundantly used codons in CCHFVs, fourteen codons [UUU (Phe), CUU (Leu), AUA (Ile), GUU (Val), UCA (Ser), CCA (Pro), ACA (Thr), GCA (Ala), CAU (His), AAA (Lys), GAA (Glu), UGU (Cys), AGA (Arg) and GGA (Gly)] had A or U at the end (A: 9; U: 5) and the remaining four (UAC, CAG, AAC, GAC) had G or C at the end. Therefore, the codons with A or U end bases are more common in the CCHFV genome consistent with previous studies (Greenbaum et al., 2008; Rabadan et al., 2006). Codon over- and under-representation analysis highlighted that RSCU values of the majority preferred and non-preferred codons lie in the range of 0.6 to 1.6. Interestingly, we found that the most over-represented codons (RSCU > 1.6) ended with A and the most under-represented (RSCU < 0.6) codons ended with G (Table 2). During the study, some under- and over-represented codons were also observed in the virus and its natural hosts. In particular, Arg (AGA, AGG) is over-represented in both CCHFVs and two potential host species (*B. taurus* and *O. aries*), but less favored in *Hyalomma*. In the case of Arg, there was a strong increase of AGA and AGG, consistent with previous studies on Influenza A virus (IAV), where AGA and AGG codons for Arg were more common than CGN (Goñi et al., 2012; Kumar et al., 2016).

Furthermore, we determined whether the CCHFV codon usage pattern is influenced by that of its hosts, including *H. sapiens*, *Hyalomma*, *B. taurus*, and *O. aries* (Table 2). We found that 45, 50, or 52 of 59 synonymous codons were equivalent between CCHFV and *H. sapiens*, *B. taurus*, or *O. aries*, respectively, while 40 of 59 synonymous codons were equivalent between CCHFV and *Hyalomma* (Table 2). Similarities in codon usage patterns between CCHFV and its natural hosts may optimize the translational efficiency of the viral genes. Specifically, Leu (UUG), Ser (AGC), Pro (CCA), Thr (ACC), Tyr (UAC), Gln (CAG), Asn (AAC), Asp (GAU, GAC), and Gly (GGC, GGA) have close homology between CCHFV and its natural hosts. Additionally, the RSCU values of several codons (Leu (UUA), Ile (AUA), Val (GUG, GUA), Ser (UCG), Thr (ACG), Ala (GCG), Arg (CGC, CGA, AGG, AGA) and Gly (GGU)) showed a strong discrepancy between CCHFV and its hosts. These results suggest that selection pressure from the hosts may be dominant on the codon usage pattern of CCHFV, which may contribute to adjusting to the cellular environment of the hosts and permit it to replicate efficiently in the hosts (Ma et al., 2015; Wong et al., 2010). Importantly, the role of selection from the hosts (*H. sapiens*, *B. taurus*, *O. aries*) in shaping codon usage patterns of CCHFV is not similar to its vector host, *Hyalomma*. Previous studies on EBOV and *Flaviviridae* members reported that the frequencies of codon usage are very different compared to codon usage patterns of hosts (Cristina et al., 2015; Schubert and Putonti, 2010). Previous studies have revealed that coincident portions of codon usage among virus and hosts may permit the corresponding amino acids to be translated efficiently, while the antagonistic portions of codon usage may enable viral proteins to be folded properly, even though the translation efficiency of the corresponding amino acids might decrease (Aragonès et al., 2010; Costafreda et al., 2014; Cristina et al., 2015; Hu et al., 2014).

3.3. Measuring the similarity effects between the overall codon usage of hosts and that of CCHFV

To further investigate codon usage similarity and to determine how the overall codon usage of the hosts and CCHFV participated in evolution, Spearman's correlational distance analysis was performed. This analysis was employed to evaluate general codon usage similarities through RSCU between CCHFV and hosts. Such RSCU-dependent analyses are performed routinely for different viral hosts, and remain restricted to codon usage and similarities (Ma et al., 2011; Wong et al., 2010; Zhou et al., 2013; Zhou et al., 2012). Here, we applied this method by performing a hierarchical clustering analysis of all species used in this study, and measured their overall codon usage similarity between virus and hosts. This newly optimized method was used to take advantage of the estimation factor and presents a clear view of codon usage (see methods Section 2.4). Two main groups were observed in this analysis. One cluster contains three mammals (*H. sapiens*, *B. taurus* and *O. aries*) and the other cluster contains the virus and the vector (*Hyalomma*) (Fig. 1). The statistical tests for distance of RSCUs (all of which were compared against a synonymous shuffling null model) suggest that a significant codon usage signature exists for vector and virus ($P < 0.01$) compared with human and virus ($P > 0.05$). This makes it clear that possible virus transmission in humans is based on the vector (*Hyalomma*) and the cycle is accomplished by vertical transmission. Models of alternative infection of arthropods and vertebrates have shown substantial constraints on arbovirus evolution (Das et al., 2013; Dieng et al., 2010; Jenkins and Holmes, 2003; Martins et al., 2012; Morrison et al., 2008; Weaver, 2006; Zhou et al., 2013). Consistent with these previous works, our findings have clearly demonstrated the overall codon usage correlation of CCHFV with *Hyalomma* and not with *H. sapiens*, *B. taurus* and *O. aries*. Based on this, we speculate that translation selection plays a major role in shaping the codon usage pattern of CCHFV.

Importantly, there are similarities between the codon usage patterns of CCHFV and its natural vector. The virus may not need a maximal expression of proteins like 'N' inside a vector, but in human or bovine/ovine hosts, all proteins including N and other enzyme proteins are important. Therefore, from an epidemic perspective, we suggest paying more attention to controlling the population of *Hyalomma* to reduce the transmission of CCHFV to humans.

3.4. General codon usage pattern in CCHFV

To estimate the magnitude of the codon usage bias within the CCHFV coding sequences, the effective number of codons (ENC) was plotted against the GC-content at the 3rd codon position (GC3). The distribution plot was used to measure the codon usage of a gene that deviates from equal usage of synonymous codons (Wright, 1990; Wu et al., 2015). In the present study, ENC values ranged from 49.95 to 55.27 (mean = 52.34 ± 1.36) (Table S1). Strikingly, all the values were higher than 35, suggesting nearly equal and slightly biased general codon usage among CCHFV genomes. Codon bias measured by ENC has the lowest value for genes with an approximate GC3 value of 0.5, however, similar ENC values can also be obtained for genes with distinct GC3 values (Wright, 1990). Our data are consistent with previous results from different RNA viruses, including CHIKV (ENC = 55.56), EBOV (ENC = 57.23), HCV (ENC = 52.62), DENV (ENC = 49.70), WNV (ENC = 53.81), CSFV (ENC = 51.7), and BODV (ENC = 50.91) (Butt et al., 2014; Cristina et al., 2015; Hu et al., 2011; Ma et al., 2013; Moratorio et al., 2013; Tao et al., 2009; Wang et al., 2011). This result suggests that viruses with slight codon bias favor effective replication in host cells with different preferences in codon usage. Previous studies suggested an inverse relationship between ENC and gene expression, given the fact that low ENC implies high gene expression and codon usage preference (Cristina et al., 2016; Nasrullah et al., 2015; Wright, 1990).

A plot of ENC values against GC3 values was constructed to check heterogeneity of codon usage (Wright, 1990). If a gene is subject to GC compositional constraints, it will lie on or near the theoretical fitting curve that represents random codon usage. In contrast, if a gene is subject to translational selection, it will lie considerably below the expected curve (Jenkins and Holmes, 2003; Wang et al., 2016). Here, the ENC value of each polyprotein-coding region of CCHFV was plotted against the corresponding GC3 content (Fig. 2). The resulting points lie considerably below the solid curve, implying that, in addition to

Table 2
The relative synonymous codon usage frequency (RSCU) of CCHFV, and its natural hosts.

AA	Codons	CCHFV	*Homo sapiens*	*Hyalomma*	*Bos taurus*	*Ovis aries*
Phe	UUU	1.07	0.97	0.65	0.87	0.94
	UUC	0.93	1.03	1.35	1.13	1.06
Leu	UUA	0.93	0.5	0.23	0.71	0.68
	UUG	1.05	0.85	0.94	1.35	1.18
	CUU	1.18	0.81	0.77	0.73	0.77
	CUC	0.75	1.07	1.42	0.93	0.99
	CUA	1.01	0.46	0.48	0.58	0.5
	CUG	1.09	2.33	2.15	1.69	1.87
Ile	AUU	1.05	1.13	0.9	0.92	0.99
	AUC	0.82	1.37	1.5	1.01	1.15
	AUA	1.13	0.5	0.6	1.07	0.87
Val	GUU	1.24	0.79	0.77	0.69	0.78
	GUC	0.93	0.9	1.25	0.82	0.91
	GUA	0.68	0.52	0.45	0.72	0.59
	GUG	1.15	1.79	1.54	1.76	1.71
Ser	UCU	1.19	1.15	0.83	0.95	1.04
	UCC	0.64	1.17	1.15	1.06	1.11
	UCA	1.48	0.93	0.86	1.4	1.24
	UCG	0.24	0.36	0.76	0.43	0.38
	AGU	1.14	0.98	0.86	0.8	0.88
	AGC	1.31	1.42	1.55	1.35	1.34
Pro	CCU	1.43	1.2	0.85	0.94	1.02
	CCC	0.78	1.22	1.14	1.01	1.12
	CCA	1.47	1.14	1.11	1.45	1.32
	CCG	0.32	0.45	0.91	0.59	0.54
Thr	ACU	1.14	1.03	0.65	0.87	0.96
	ACC	1	1.32	1.22	1.09	1.17
	ACA	1.64	1.19	1.13	1.44	1.33
	ACG	0.22	0.46	1	0.6	0.55
Ala	GCU	1.16	1.08	1.16	0.97	1.03
	GCC	0.85	1.51	1.23	1.13	1.31
	GCA	1.85	0.95	1	1.3	1.16
	GCG	0.14	0.46	0.6	0.6	0.51
Tyr	UAU	0.94	0.93	0.54	0.9	0.87
	UAC	1.06	1.07	1.46	1.1	1.13
His	CAU	1.04	0.85	0.82	0.88	0.86
	CAC	0.96	1.15	1.18	1.12	1.14
Gln	CAA	0.95	0.49	0.85	0.71	0.65
	CAG	1.05	1.51	1.15	1.29	1.35
Asn	AAU	0.88	0.98	0.59	0.87	0.92
	AAC	1.12	1.02	1.41	1.13	1.08
Lys	AAA	1.05	0.88	0.73	0.89	0.94
	AAG	0.95	1.12	1.27	1.11	1.06
Asp	GAU	0.99	0.99	0.76	0.85	0.93
	GAC	1.01	1.01	1.24	1.15	1.07
Glu	GAA	1.12	0.85	0.87	0.92	0.96
	GAG	0.88	1.15	1.13	1.08	1.04

Table 2 (*continued*)

Cys	UGU	1.03	0.95	0.72	0.78	0.84
	UGC	0.97	1.05	1.28	1.22	1.16
Arg	CGU	0.28	0.54	0.68	0.26	0.3
	CGC	0.19	1.11	1.36	0.52	0.57
	CGA	0.33	0.76	1.23	0.27	0.39
	CGG	0.21	1.31	1	0.73	0.79
	AGA	2.78	1.18	0.84	2.16	2.04
	AGG	2.21	1.1	0.89	2.07	1.91
Gly	GGU	1.1	0.71	0.79	0.51	0.6
	GGC	1.06	1.35	1.37	1.01	1.09
	GGA	1.11	1.01	1.27	1.25	1.2
	GGG	0.74	0.93	0.57	1.23	1.12

AA represents amino acid; the "RSCU" value represents the pattern of relative synonymous codon usage; orange colors represents the codons favored by CCHFV and hosts (RSCU > 1); over-represented (RSCU > 1.6), and under-represented (RSCU < 0.6) codons are marked as bold with red and green colors, respectively; the ideal codons for CCHFV are marked as underline.

mutation pressure, other factors, such as translational selection, also influence the codon usage pattern of CCHFV. This result is consistent with related plots in prior studies (Butt et al., 2014; Chen et al., 2014; Wang et al., 2016).

To evaluate the factors affecting CCHFV codon usage bias, a neutrality plot was constructed between GC12 and GC3 in order to determine the influence of mutation bias and natural selection. In the plot, the regression slope estimates the degree of neutrality and selects the effect that influences evolution (Nasrullah et al., 2015). If the correlation between GC12 and GC3 is significant, then mutation pressure is the main force shaping codon usage bias. The neutrality plot showed that there was no significant correlation between GC12 and GC3 ($r = 0.422$, $P > 0.01$), suggesting that both natural selection and mutation pressure influence the codon usage pattern of CCHFV (Fig. S3). Our result is inconsistent with a previous study showing that mutation pressure had a stronger influence in shaping the codon usage pattern of *Paeonia lactiflora* (Wu et al., 2015).

3.5. Relationship between relative abundance of dinucleotides and codon usage in CCHFV

Previous studies have suggested that the dinucleotide compositional constraints of the genome are potentially involved in codon usage bias (Karniychuk, 2016; Kumar et al., 2016; Liu et al., 2012; Wang et al., 2016). Therefore, we calculated the relative abundance of 16 dinucleotides from the complete coding sequences of CCHFV. The result showed that the occurrences of dinucleotides in CCHFV are not random, and also that no dinucleotide exists at the expected frequencies. In particular, UG and CA were over-represented ($\rho xy > 1.23$) whereas UA and CG were under-represented ($\rho xy < 0.78$) (Table 3). This result is similar to previous studies, which found that both UA and CG are under-represented in different RNA viruses (Burge et al., 1992; Butt et al., 2014; Wang et al., 2016). Furthermore, the RSCU values of the eight codons containing CG (CCG, CGC, ACG, GCG, CGU, CGC, CGA, and UCG) and the codons containing UA, suggests that all these codons are not preferentially selected, except for UAC. Altogether, our results suggest that dinucleotide composition plays a role in the synonymous codon usage pattern of CCHFV.

We also found that the relatively low abundances of CpG and UpA in CCHFV may be helpful for the virus to escape the host antiviral immune response and complete virus transcription (Kumar et al., 2016; Wang et al., 2016). Unmethylated CpG can be recognized by the host innate
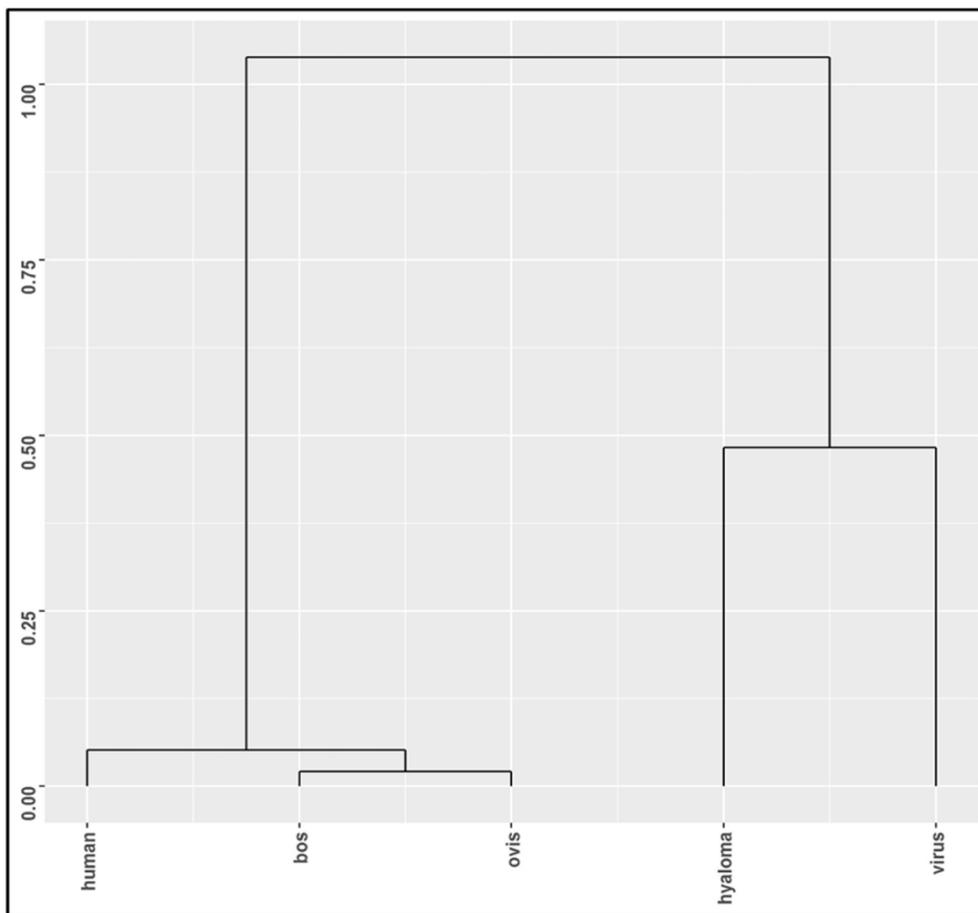
Fig. 1. Similarity distance analysis of the codon usage using CCHFV and its hosts (Spearman correlational distances = 1 − SpearmanRho).
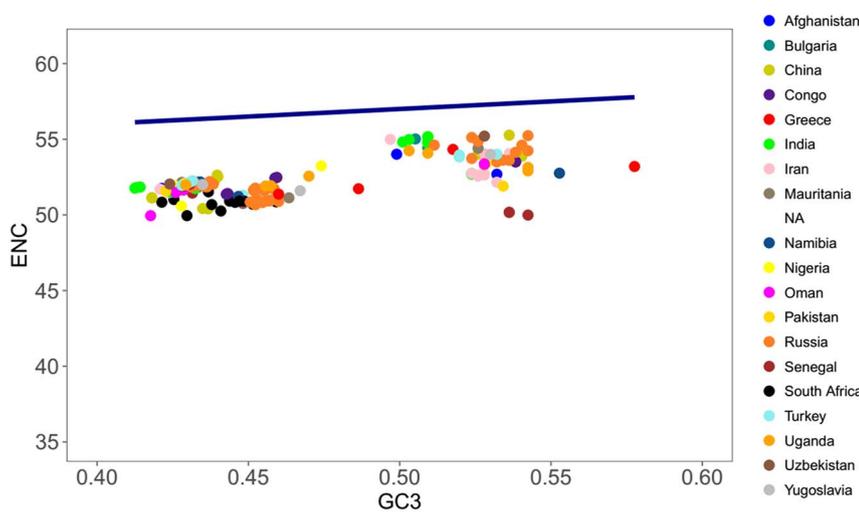
Fig. 2. Codon ENC analysis of CCHFV genomes. The effective number of codons (ENC-values, Y-axis) was plotted against the GC-content at the third synonymous codon positions (GC3-values, X-axis). The continuous blue line represents theoretical ENC values for random codon usage as a function of GC3. Deviation from this line in the direction of lower ENC values points to the selection of a preferred set of synonymous codons. Some CCHFV genes are far away from the standard line, showing that their codon usage pattern might be affected by other factors besides nucleotide composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

immune system as a pathogen signature, and can activate various immune response pathways (Cheng et al., 2013; Greenbaum et al., 2008; Shackelton et al., 2006; Wang et al., 2016). Recognition of unmethylated CG by Toll like receptor 9 (TLR9), leads to activation of several immune response pathways (Dorn and Kippenberger, 2008). The vertebrate immune system relies on unmethylated CG recognition in DNA molecules as a signature of infection, and CG under-representation in RNA viruses is exclusively observed in vertebrate viruses; therefore, it is reasonable to suggest that a TLR9-like mechanism exists in the vertebrate immune system that recognizes CG when in an RNA context (such as in the genomes of RNA viruses) and

triggers immune responses (Pena et al., 2009). The deficiency of UA was proposed to help viruses by reducing the risk of nonsense mutations, minimizing improper transcription and decreasing the opportunities of cleavage by RNase L (Al-saif and Khabar, 2012; Duan et al., 2015).

3.6. Variation in codon usage among CCHFVs

Correspondence analysis (COA) was performed to determine the trend in codon usage variation among the coding sequences of different CCHFV strains. The analysis is used to identify the systematic

Table 3
Relative abundance of the 16 dinucleotides in polyprotein-coding region of 179 CCHFVs strains.

	AA	AC	AG	AT	CA	CC	CG	CT
Mean ± STD	1.06 ± 0.06	0.98 ± 0.05	1.14 ± 0.08	0.82 ± 0.06	1.33 ± 0.06	0.97 ± 0.07	0.31 ± 0.05	1.21 ± 0.08
Range	0.98–1.17	0.84–1.06	0.98–1.24	0.75–0.93	1.23–1.45	0.83–1.11	0.23–0.44	1.11–1.38
	GA	GC	GG	GT	TA	TC	TG	TT
Mean ± STD	1.03 ± 0.06	1.04 ± 0.04	1.01 ± 0.05	0.91 ± 0.05	0.58 ± 0.12	1.02 ± 0.12	1.42 ± 0.09	1.12 ± 0.04
Range	0.91–1.14	0.92–1.11	0.88–1.14	0.79–1.00	0.36–0.72	0.90–1.21	1.27–1.57	0.98–1.23

Ratios of observed dinucleotide frequencies to expected dinucleotide frequencies that are further normalized by considering the amino acid usage of the nucleotide sequence.

relationships among variables. Additionally, it simplifies complex data to deliver different strains or genes in multidimensional space (Butt et al., 2014; Greenacre, 1984; Kumar et al., 2016). The COA was carried out on the relative synonymous codon usage (RSCU) values for each segment strain (S, M and L) of CCHFV, and was used to determine allocation in the first two principal axes of the plane. Both principal axes accounted for the following percentages of total variation: 51.62% and 22.01% in the S segment; 72.96% and 15.66% in the M segment: and 50.46% and 25.7% in the L segment (Fig. 3). These results suggest that the second axis represents the countries where the virus arises, and the first axis represents the virus strains (see below).

Scattered data in the principal axis represents different geographical lineages and their relationship with each other. The COA (S, M, and L segments) showed that all CCHFV isolates were assembled into clusters (Fig. 4). All Russian, Turkish, Yugoslavian, and Bulgarian CCHFV isolates were present in one cluster, whereas CCHFV isolates from Pakistan, Afghanistan, Oman, and Iran were found in another cluster. Isolates from Congo (the site of the first outbreak of CCHFV) and Uganda were found in one cluster while some Chinese isolates were found in a cluster with India. Isolates from Nigeria were distributed in different clusters. Interestingly, the same genetic lineage was found between Congo-Uganda, and Russia-Turkey, where both relatives were isolated from the same origin (Fig. 4). These analyses indicate that the geographical spots play a key role in CCHFV evolution and in a pattern of synonymous codon usage, and such analyses may help to trace the root of emerging CCHFV strains in the future. Additionally, current results also highlight that every infected country has more than one prevalent genetic lineage.

CCHFV is the most genetically diverse of the arboviruses, with nucleotide sequence differences among isolates ranging from 20% for the viral S segment to 31% for the M segment. Viruses with diverse sequences can be found within the same geographic area, while closely related viruses have been isolated in far distant regions (Bente et al., 2013), suggesting that widespread dispersion of CCHFV has occurred in the past, possibly by ticks carried on migratory birds or through the international livestock trade. Moreover, re-assortment among genome segments during co-infection of ticks and vertebrates appears to have played an important role in generating diversity, and represents a potential future source of novel viruses (Bente et al., 2013).

A phylogenetic analysis was performed using the maximum likelihood method in order to evaluate the effect of evolutionary processes on the CCHFV codon usage pattern. Similar to the pattern observed in COA (Fig. 4), all CCHFV isolates were grouped into separate clades (Fig. 5). This supports the dominant effect of evolutionary processes and geographical distribution on codon usage patterns. The phylogenetic analysis also revealed evidence of genome re-assortment and recombination during co-infection of a single host, indicating the potential for the future emergence of novel variants (Chamberlain et al., 2005; Hewson et al., 2004).

3.7. Codon usage adaptation in CCHFV

Codon adaptation index (CAI) analyses were performed to determine the codon usage optimization and adaptation of CCHFV in relation to its hosts. CAI values for all codons were calculated by taking the codon usage of *H. sapiens*, *B. taurus*, *O. aries*, and *Hyalomma* as a reference set. The CAI values range from 0 to 1, and higher CAI values signify higher levels of codon usage bias (Butt et al., 2014). This study found a tendency for higher CAI values (> 0.5), representing adaptability of codon usage of CCHFV to its hosts, with the consequence of lower translation efficiency. We found that, in relation to *H. sapiens*, *B. taurus*, *O. aries*, and *Hyalomma*, the CAI values of CCHFV polyprotein-coding regions were 0.80 ± 0.02, 0.73 ± 0.02, 0.78 ± 0.02 and 0.64 ± 0.03. The Student's *t*-test was performed to measure the significant differences in the data, and it indicated statistically significant differences in CAI values (Table S2).
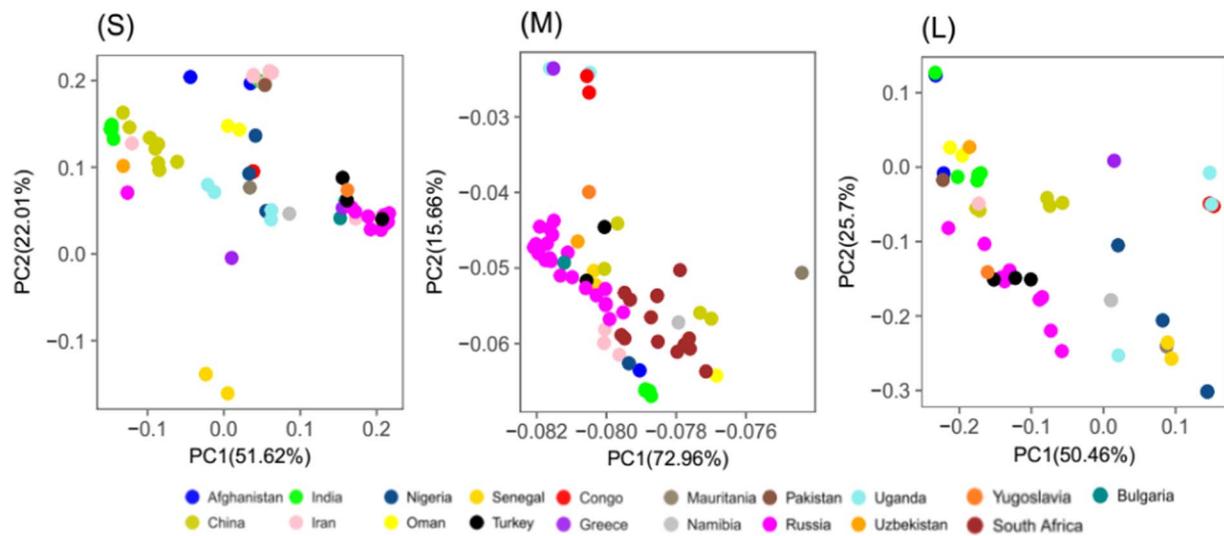
Fig. 3. A plot of values of the first axis (Axis 1) and the second axis (Axis 2) of each polyprotein-coding region of CCHFV in correspondence analysis. The first axis accounts for 51.62%, 72.96%, and 50.46% of total variation, and the second axis accounts for 22.01%, 15.66%, and 25.7% of total variation. S: small segment, M: medium segment, L: large segment.
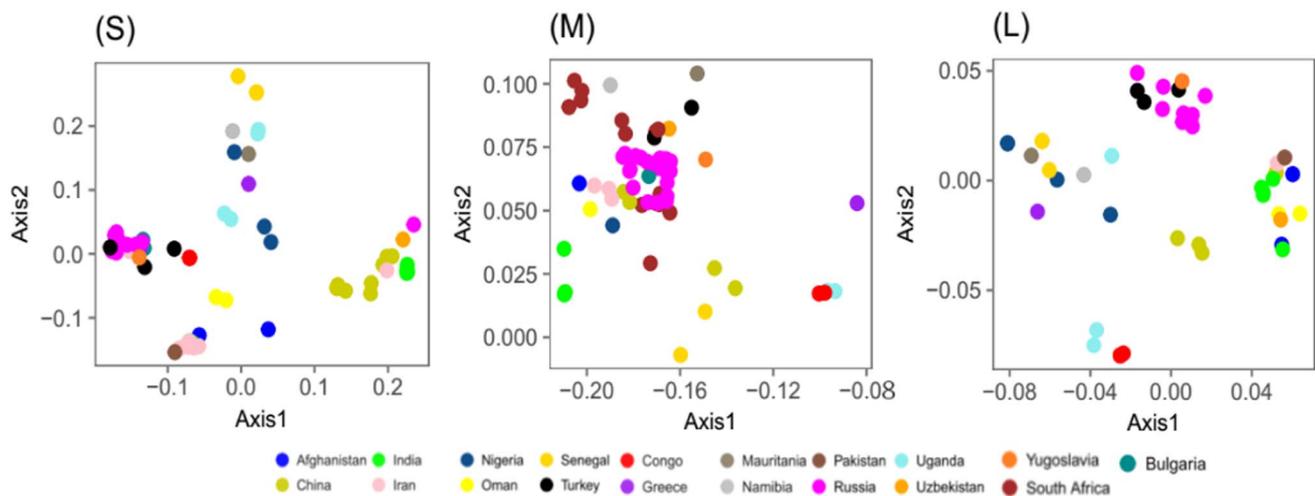
Fig. 4. The correspondence analysis (COA) of the genes in CCHFV genomes. Each point represents a gene corresponding to the coordinates of the first and second axes of variation generated from the correspondence analysis. S: small segment, M: medium segment, L: large segment.

To further validate the observed statistically significant differences in CAI values (Puigbò et al., 2008a), the expected CAI (e-CAI) values were computed for CCHFV coding sequences in relation to *H. sapiens*, *B. taurus*, *O. aries*, and *Hyalomma* codon usage sets. The e-CAI is an executable program (Puigbò et al., 2008b) that measures the expected value of the CAI by generating 500 uneven sequences with similar nucleotide contents and amino acid composition as the sequences of interest. A Kolmogorov–Smirnov test was applied to determine the e-CAI of these random sequences and to confirm that the generated random sequences had a normal distribution. The e-CAI values of 0.795 ($P < 0.05$), 0.737 ($P < 0.05$), 0.782 ($P < 0.05$), and 0.658 ($P < 0.05$) for *H. sapiens*, *B. taurus*, *O. aries* and *Hyalomma*, respectively, revealed that the generated sequences had a normal distribution.

Altogether the results indicated that the CAI values for CCHFVs with reference to *H. sapiens*, *B. taurus*, and *O. aries* are significantly distinct from the CAI values acquired from *Hyalomma* (Fig. S4). The tendency for higher CAI values for *H. sapiens*, *B. taurus* and *O. aries* suggests that selection pressure from *H. sapiens*, *B. taurus*, and *O. aries* can affect the codon usage of CCHFV and that the evolution of codon usage in CCHFV has allowed it to use the translation resource of *H. sapiens*, *B. taurus*, and *O. aries* more efficiently. In addition, the results suggest that these differences are related to codon usage preferences. Our results about

codon usage preferences are consistent with published results that have suggested dissimilar patterns between EBOV and human genes, and between ZKV and human, *A. aegypti*, and *A. albopictus* genes (Cristina et al., 2016; Cristina et al., 2015).

3.8. Influencing factors of codon usage pattern

We considered two determinants, mutation bias and natural selection, to examine codon usage bias in CCHFV. For this purpose, a correlation analysis was performed between CAI and ENC values. If the correlation (r) between the two indices approaches 1, this suggests that translational selection is preferred over mutation. Otherwise, if the r value approaches 0 (no correlation), mutation may be more influential than translational selection (Wang et al., 2016). A positive correlation was observed among the CAI values of CCHFV genes in relation to *H. sapiens* ($r = 0.41$, $P < 0.01$), *B. taurus* ($r = 0.42$, $P < 0.01$), *O. aries* ($r = 0.43$, $P < 0.01$), and *Hyalomma* ($r = 0.64$, $P < 0.01$), with ENC values that suggested the influence of both translation selection and mutational pressure on the codon usage pattern of CCHFV (Table 4). Such influence can also be analyzed using Spearman's rank correlation among GC3s (0.55), GC contents (0.50), hydrophobicity (-0.70), and aromaticity (0.81) with ENC (Table 5). Among these GC3 and GC
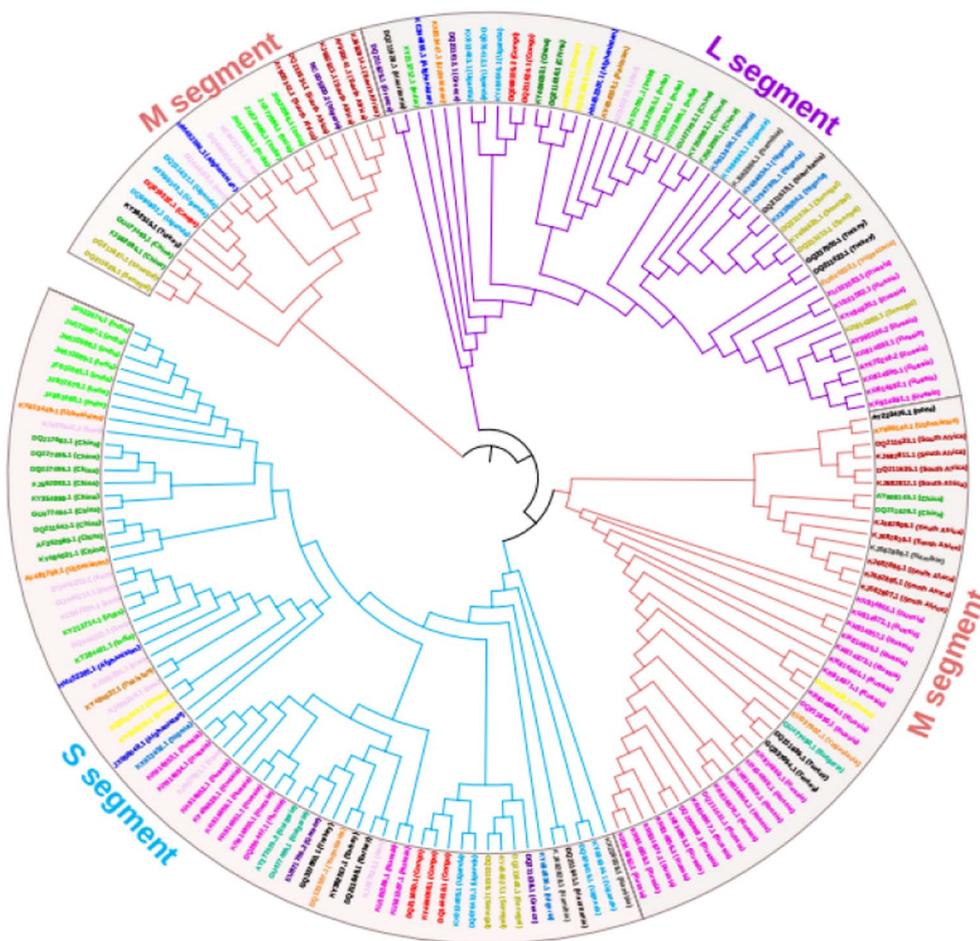
Fig. 5. Phylogenetic tree based on the polyprotein-coding regions of 179 CCHFV strains (S: small segment, M: medium segment, L: large segment). The tree was generated by the maximum likelihood (ML) method using Clustal × 2. The tree was designed by using the online tool "iTOL".

Table 4
The correlation analysis between CAI and ENC.

	CAI (*H. sapiens*)	CAI (*B. taurus*)	CAI (*O. aries*)	CAI (*Hyalomma*)
ENC	0.42*	0.42*	0.41*	0.46*

The numbers in each column represent correlation coefficient "r" values, which are calculated in each correlation analysis.
NS non-significant ($P > 0.05$).
 * Represents $P < 0.01$.

Table 5
Correlation analysis among GC, GC3, GRAVY, AROMO, ENC, and the first two principal axes in the polyprotein-coding region of CCHFV isolates.

Variables	GC	GC3	AROMO	GRAVY	ENC	Axis1	Axis2
GC		0.95*	0.58*	− 0.55*	0.50*	0.62*	0.30*
GC3			0.64*	− 0.65*	0.55*	0.73*	0.20*
AROMO				− 0.91*	0.81*	0.87*	− 0.22*
GRAVY					− 0.70*	− 0.83*	0.22*
ENC						0.77*	− 0.27*
Axis1							− 0.16^NS

The numbers in each column represent correlation coefficient "r" values, which are calculated in each correlation analysis.
Gravy general average hydrophobicity; ARO aromaticity.
NS non-significant ($P > 0.05$).
 * Represents $P < 0.01$.

contents, a significant correlation was observed, supporting the result and signifying their role in altering CCHFV codon usage patterns (Fig. 6). This result reflects the influence of translational selection and mutational pressure on the codon usage pattern of CCHFV. Previous

studies suggested that natural selection was generally determined by the base contents at the first and second positions, while mutational pressure is mostly determined by the base contents at the third codon positions (Hu et al., 2014; Roychoudhury and Mukherjee, 2010).

To confirm whether translation selection from the hosts plays a role in shaping the codon usage pattern of CCHFV, the tAI values were calculated based on the tRNA copy numbers of *H. sapiens*. The results indicated that the tAI values of CCHFV strains range from 0.290 to 0.303, with an average value of 0.296 and a SD of 0.003. Moreover, the negative correlation between tAI and CAI values ($r = − 0.547$, $P < 0.01$) in CCHFV highlights the importance of translational selection in the formation of synonymous codon usage pattern.

Hydrophobicity (GRAVY), and aromaticity (AROMO) may also be related to the codon usage pattern of viruses (Wang et al., 2016). It is obvious from the table that the GRAVY values have significant negative correlations with GC ($r = − 0.54$, $P < 0.01$), GC3s ($r = − 0.83$, $P < 0.01$), and ENC ($r = − 0.77$, $P < 0.01$). Conversely, AROMO values had a significant positive correlation with GC ($r = 0.62$, $P < 0.01$), GC3s ($r = 0.87$, $P < 0.01$), and ENC ($r = 0.81$, $P < 0.01$) (Table 5). The results indicate that the degree of protein hydrophobicity and aromaticity are associated with the codon usage variation in CCHFV, highlighting the importance of natural translational selection on forming the codon usage pattern (Chen et al., 2014). The involvement of aromaticity and hydrophobicity in the construction of codon usage bias has been revealed in some RNA viruses, such as classical swine fever virus, and duck hepatitis A virus (Chen et al., 2014; Tao et al., 2009). Furthermore, there was a significant negative correlation of GRAVY values with axis 1, and a significant positive correlation of GRAVY with axis 2. In contrast, there was a significant positive correlation of AROMO with axis 1, and a significant negative correlation of
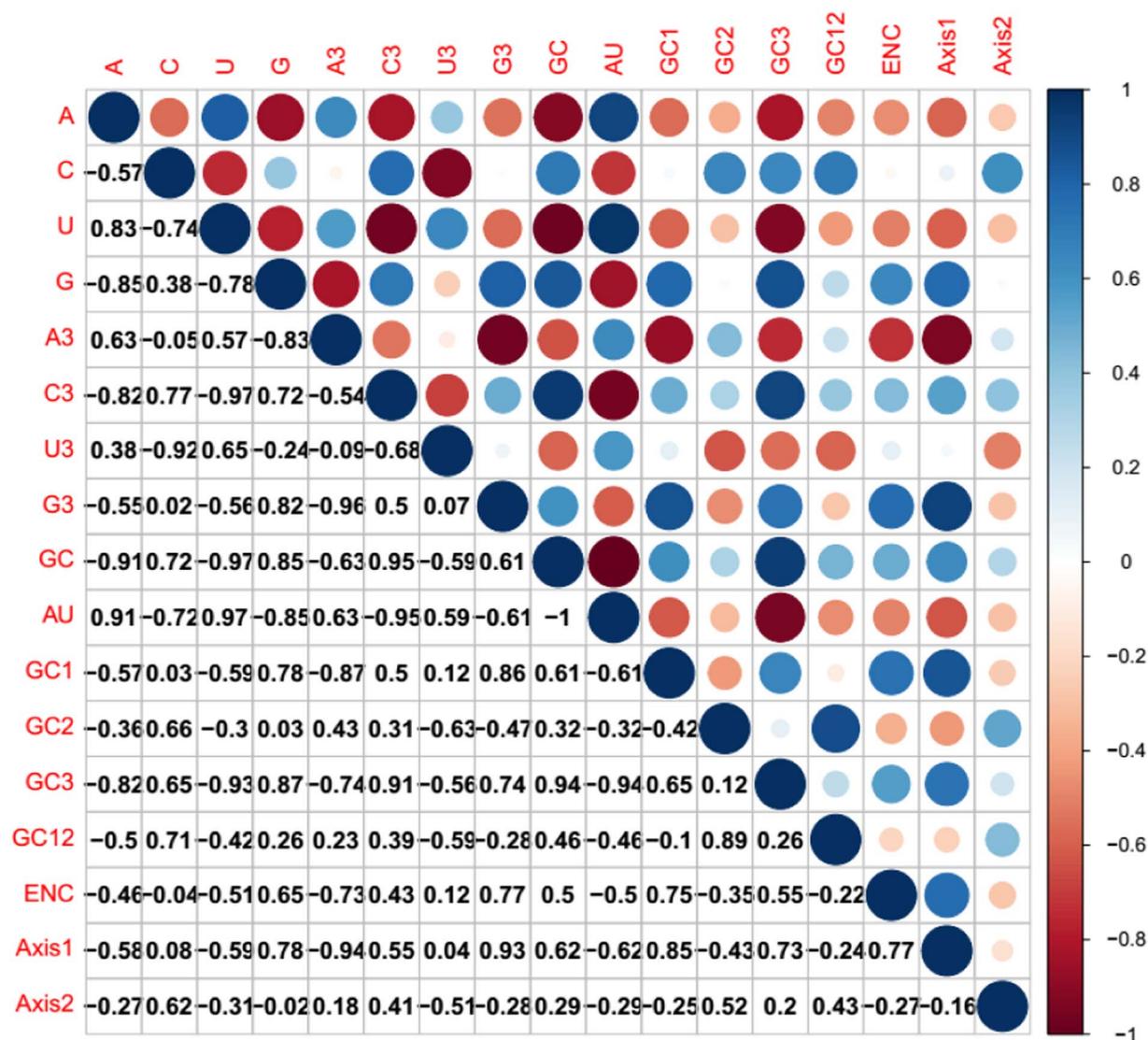
Fig. 6. Correlation analysis among different nucleotides compositions of CCHFVs, and the first two principal axes of COA. The dark blue means positive correlation, and the dark red means negative correlation; the value larger means more significant correlation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

AROMO with axis 2. These results implied that both Axis 1 and Axis 2 have significant roles in shaping the CCHFV codon usage pattern (Table 5), and additionally suggest that the aromaticity and hydrophobicity of proteins are related to the codon usage pattern of CCHFV. Aromaticity and hydrophobicity are known to play a role in peptide self-assembly and protein aggregation rates (Wang et al., 2016).

Spearman's rank correlation analysis was also performed among the nucleotide contents i.e., A, U, C, G, GC and A3, U3, C3, G3, GC3s. A significant correlation was seen among A3, A, and G. U3 also showed a significant correlation with U, C, and GC. C3 had significant correlations with A, U, C, and GC. G3 had a significant correlation with A, U, and G. Furthermore, GC3 had a significant correlation with A, U, G, and GC (Fig. 6). This analysis showed the influence of nucleotide contents on codon usage pattern.

The correlation analysis was also performed between the first two-principle axes (axis 1 and axis 2) and nucleotide contents of CCHFV genomes. Results showed various significant correlations between the two principle axes and nucleotide contents (Fig. 6). The first axis exhibited a statistically significant correlation with U3 ($r = 0.04$, $P < 0.001$), G3 ($r = 0.93$, $P < 0.001$), and GC1 ($r = 0.85$, $P < 0.001$), while the second axis had a significant correlation with C3 ($r = 0.41$, $P < 0.001$), A3 ($r = 0.18$, $P < 0.001$), G ($r = -0.02$,

$P < 0.001$), and GC ($r = 0.29$, $P < 0.001$). These results suggest that nucleotide contents influence the pattern of synonymous codon usage.

4. Conclusion

This study demonstrated that the codon usage bias of CCHFV is weak and that, in addition to translation selection, mutation pressure also influences the codon usage bias. Other factors, such as base composition, aromaticity, and hydrophobicity, also have an effect on the codon usage pattern. Importantly, there are similarities between the codon usage patterns of CCHFV and its natural hosts. Further studies will be required to establish viral adaptation in various aspects and hosts that will help researchers understand and control of CCHFV infection and transmission.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2017.11.027.

Author contributions

SUR, DKC and SHT conceived and designed experiments; SUR and XTY performed all experiments. SUR, XCL and XTY design and analyzed the data. SUR and XTY drafted the manuscript. All authors read and

approved the final manuscript.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

We would like to thank the members of the Bioinformatics Center of Northwest A&F University for their useful input.

References

Al-saif, M., Khabar, K.S.A., 2012. UU/UA dinucleotide frequency reduction in coding regions results in increased mRNA stability and protein expression. Mol. Ther. 20, 954–959. http://dx.doi.org/10.1038/mt.2012.29.

Andersson, Siv G.E., Kurland, C.G., 1990. Codon preferences in free-living microorganisms. Microbiol. Rev. 54, 198–210.

Aragonès, L., Guix, S., Ribes, E., Bosch, A., Pintó, R.M., 2010. Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. PLoS Pathog. 6, 1–12. http://dx.doi.org/10.1371/journal.ppat.1000797.

Aslam, S., Latif, M.S., Daud, M., Rahman, Z.I.A.U.R., Tabassum, B., Riaz, M.S., Khan, A., Tariq, M., Husnain, T., 2016. Crimean-Congo hemorrhagic fever: risk factors and control measures for the infection abatement (review) 15–20. doi:https://doi.org/10.3892/br.2015.545.

Athey, J., Alexaki, A., Osipova, E., Rostovtsev, A., Santana-Quintero, L.V., Katneni, U., Simonyan, V., Kimchi-Sarfaty, C., 2017. A new and updated resource for codon usage tables. BMC Bioinformatics 18, 391. http://dx.doi.org/10.1186/s12859-017-1793-7.

Bente, D.A., Forrester, N.L., Watts, D.M., Mcauley, A.J., Whitehouse, C.A., Bray, M., 2013. Crimean-Congo hemorrhagic fever: history, epidemiology, pathogenesis, clinical syndrome and genetic diversity. Antivir. Res. 100, 159–189. http://dx.doi.org/10.1016/j.antiviral.2013.07.006.

Bertolotti-ciarlet, A., Smith, J., Strecker, K., Paragas, J., Altamura, L.A., Mcfalls, J.M., Frias-sta, N., Schmaljohn, C.S., Doms, R.W., 2005. Cellular localization and antigenic characterization of Crimean-Congo hemorrhagic fever virus glycoproteins. Society 79, 6152–6161. http://dx.doi.org/10.1128/JVI.79.10.6152.

Burge, C., Campbell, A.M., Karlin, S., 1992. Over- and under-representation of short oligonucleotides in DNA sequences. Proc. Natl. Acad. Sci. 89, 1358–1362. http://dx.doi.org/10.1073/pnas.89.4.1358.

Burns, C.C., Shaw, J., Campagnoli, R., Jorba, J., Vincent, A., Quay, J., Kew, O., Irol, J.V, 2006. Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region 80, 3259–3272. doi:https://doi.org/10.1128/JVI.80.7.3259.

Butenko, A.M., Karganova, G.G., 2007. Crimean-Congo hemorrhagic fever in Russia and other countries of the former Soviet Union. In: Ergonul, O., Whitehouse, C.A. (Eds.), Crimean-Congo Hemorrhagic Fever: A Global Perspective. Springer Netherlands, Dordrecht, pp. 99–114. http://dx.doi.org/10.1007/978-1-4020-6106-6_9.

Butt, A.M., Nasrullah, I., Tong, Y., 2014. Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. PLoS One 9. http://dx.doi.org/10.1371/journal.pone.0090905.

Butt, A.M., Nasrullah, I., Qamar, R., Tong, Y., 2016. Evolution of codon usage in Zika virus genomes is host and vector specific. Emerg. Microbes Infect. 1–14. http://dx.doi.org/10.1038/emi.2016.106.

Chamberlain, J., Cook, N., Lloyd, G., Mioulet, V., Tolley, H., Hewson, R., 2005. Co-evolutionary patterns of variation in small and large RNA segments of Crimean-Congo hemorrhagic fever virus. J. Gen. Virol. 86, 3337–3341. http://dx.doi.org/10.1099/vir.0.81213-0.

Chen, H., Sun, S., Norenburg, J.L., Sundberg, P., 2014. Mutation and selection cause codon usage and bias in mitochondrial genomes of ribbon worms (Nemertea). PLoS One 9. http://dx.doi.org/10.1371/journal.pone.0085631.

Cheng, X., Virk, N., Chen, W., Ji, S., Ji, S., Sun, Y., Wu, X., 2013. CpG usage in RNA viruses: data and hypotheses. PLoS One 8. http://dx.doi.org/10.1371/journal.pone.0074109.

Comeron, J.M., Aguade, M., 1998. An evaluation of measures of synonymous codon usage bias. J. Mol. Evol. 47, 268–274. http://dx.doi.org/10.1007/pl00006384.

Costafreda, M.I., Pérez-Rodriguez, F.J., D'Andrea, L., Guix, S., Ribes, E., Bosch, A., Pintó, R.M., 2014. Hepatitis A virus adaptation to cellular shutoff is driven by dynamic adjustments of codon usage and results in the selection of populations with altered capsids. J. Virol. 88, 5029–5041. http://dx.doi.org/10.1128/JVI.00087-14.

Cristina, J., Moreno, P., Moratorio, G., Musto, H., 2015. Genome-wide analysis of codon usage bias in ebolavirus. Virus Res. 196, 87–93. http://dx.doi.org/10.1016/j.virusres.2014.11.005.

Cristina, J., Fajardo, A., So, M., Moratorio, G., Musto, H., 2016. A detailed comparative analysis of codon usage bias in Zika virus 223, 147–152. doi:https://doi.org/10.1016/j.virusres.2016.06.022.

Das, B., Das, M., Dwibedi, B., Kar, S.K., Hazra, R.K., 2013. Molecular investigations of dengue virus during outbreaks in Orissa state, Eastern India from 2010 to 2011. Infect. Genet. Evol. 16, 401–410. http://dx.doi.org/10.1016/j.meegid.2013.03.016.

Dieng, H., Saifur, R.G.M., Hassan, A.A., Che Salmah, M.R., Boots, M., Satho, T., Jaal, Z., Abubakar, S., 2010. Indoor-breeding of *Aedes albopictus* in northern peninsular Malaysia and its potential epidemiological implications. PLoS One 5. http://dx.doi.org/10.1371/journal.pone.0011790.

Dorn, A., Kippenberger, S., 2008. Clinical application of CpG-, non-CpG-, and antisense oligodeoxynucleotides as immunomodulators. Curr. Opin. Mol. Ther. 10, 10–20.

Duan, X., Yi, S., Guo, X., Wang, W., 2015. A comprehensive analysis of codon usage patterns in blunt snout bream (*Megalobrama amblycephala*) based on RNA-Seq data. Int. J. Mol. Sci. 3, 11996–12013. http://dx.doi.org/10.3390/ijms160611996.

Elliott, R.M., 2017. Review Article Molecular Biology of the Bunyaviridae. pp. 501–522.

European Centre for Disease Prevention and Control. Factsheet for health professionals: Crimean-Congo haemorrhagic fever (CCHF) 2016. Available from: http://ecdc.europa.eu/en/healthtopics/emerging_and_vectorborne_diseases/tick_borne_diseases/crimean_congo/factsheet-health-professionals/Pages/factsheet.aspx.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y.H., Zhang, J., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80. http://dx.doi.org/10.1186/gb-2004-5-10-r80.

Ghosh, S., Nagar, G., 2014. Problem of ticks and tick-borne diseases in India with special emphasis on progress in tick control research: a review. J. Vector Borne Dis. 51, 259–270.

Goñi, N., Iriarte, A., Comas, V., Soñora, M., Moreno, P., Moratorio, G., Musto, H., Cristina, J., 2012. Pandemic influenza A virus codon usage revisited: biases, adaptation and implications for vaccine strain development. Virol. J. 9, 263. http://dx.doi.org/10.1186/1743-422X-9-263.

Greenacre, M.J., 1984. Theory and Applications of Correspondence Analysis.

Greenbaum, B.D., Levine, A.J., Bhanot, G., Rabadan, R., 2008. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. PLoS Pathog. 4, 1–9. http://dx.doi.org/10.1371/journal.ppat.1000079.

van Hemert, F., Berkhout, B., 2016. Nucleotide composition of the Zika virus RNA genome and its codon usage. Virol. J. 13, 95. http://dx.doi.org/10.1186/s12985-016-0551-1.

van Hemert, F., van der Kuyl, A.C., Berkhout, B., 2016. Impact of the biased nucleotide composition of viral RNA genomes on RNA structure and codon usage. J. Gen. Virol. 97, 2608–2619. http://dx.doi.org/10.1099/jgv.0.000579.

Hershberg, R., Petrov, D.A., 2008. Selection on codon bias. TL - 42. Annu. Rev. Genet. 42, 287–299. http://dx.doi.org/10.1146/annurev.genet.42.110807.091442.

Hewson, R., Gmyl, A., Gmyl, L., Smirnova, S.E., Karganova, G., Jamil, B., Hasan, R., Chamberlain, J., Clegg, C., 2004. Evidence of segment reassortment in Crimean-Congo haemorrhagic fever virus. J. Gen. Virol. 85, 3059–3070. http://dx.doi.org/10.1099/vir.0.80121-0.

Hoogstraal, H., 1979. The epidemiology of tick-borne Crimean-Congo hemorrhagic fever in Asia, Europe, and Africa. J. Med. Entomol. 15, 307–417.

Hu, J., Wang, Q., Zhang, J., Chen, H., Xu, Z., Zhu, L., Ding, Y., Ma, L., Xu, K., Gu, Y., Liu, Y., 2011. The characteristic of codon usage pattern and its evolution of hepatitis C virus. Infect. Genet. Evol. 11, 2098–2102. http://dx.doi.org/10.1016/j.meegid.2011.08.025.

Hu, C., Chen, J., Ye, L., Chen, R., Zhang, L., Xue, X., 2014. Codon usage bias in human cytomegalovirus and its biological implication. Gene 545, 5–14. http://dx.doi.org/10.1016/j.gene.2014.05.018.

Jenkins, G.M., Holmes, E.C., 2003. The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res. 92, 1–7 10.1016/S0168-1702(02)00309-X.

Jung, S.-K., McDonald, K., 2011. Visual gene developer: a fully programmable bioinformatics software for synthetic gene optimization. BMC Bioinformatics 12, 340. http://dx.doi.org/10.1186/1471-2105-12-340.

Karniychuk, U.U., 2016. Analysis of the synonymous codon usage bias in recently emerged enterovirus D68 strains. Virus Res. 223, 73–79. http://dx.doi.org/10.1016/j.virusres.2016.06.015.

Kattoor, J.J., Malik, Y.S., Sasidharan, A., Rajan, V.M., Dhama, K., Ghosh, S., Banyai, K., Kobayashi, N., Singh, R.K., 2015. Analysis of codon usage pattern evolution in avian rotaviruses and their preferred host. Infect. Genet. Evol. 34, 17–25. http://dx.doi.org/10.1016/j.meegid.2015.06.018.

Kumar, N., Bera, B.C., Greenbaum, B.D., Bhatia, S., Sood, R., Selvaraj, P., Anand, T., Tripathi, B.N., Virmani, N., 2016. Revelation of influencing factors in overall codon usage bias of equine influenza viruses. PLoS One 11, e0154376. http://dx.doi.org/10.1371/journal.pone.0154376.

Lara-Ramírez, E.E., Salazar, M.I., López-López, M.D.J., Salas-Benito, J.S., Sánchez-Varela, A., Guo, X., 2014. Large-scale genomic analysis of codon usage in dengue virus and evaluation of its phylogenetic dependence. Biomed. Res. Int. 2014, 1–9. http://dx.doi.org/10.1155/2014/851425.

Leblebicioglu, H., 2010. International Journal of Antimicrobial Agents Crimean – Congo Haemorrhagic Fever in Eurasia 2000, 43–46. http://dx.doi.org/10.1016/j.ijantimicag.2010.06.020.

Leblebicioglu, H., Ozaras, R., Irmak, H., Sencan, I., 2016. Crimean-Congo hemorrhagic fever in Turkey: Current status and future challenges. Antiviral Res. 126, 21–34. http://dx.doi.org/10.1016/j.antiviral.2015.12.003.

Letunic, I., Bork, P., 2011. Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 39, 475–478. http://dx.doi.org/10.1093/nar/gkr201.

Liu, X., Zhang, Y., Fang, Y., Wang, Y., 2012. Patterns and influencing factor of synonymous codon usage in porcine circovirus. Virol. J. 9, 68. http://dx.doi.org/10.1186/

1743-422X-9-68.

Liu, H., Rahman, S.U., Mao, Y., Xu, X., Tao, S., 2017. Codon usage bias in 5′ terminal coding sequences reveals distinct enrichment of gene functions. Genomics 0–1. http://dx.doi.org/10.1016/j.ygeno.2017.07.008.

Ma, M., Ha, X., Ling, H., Wang, M., Zhang, F., Zhang, S., 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. pp. 1–10. http://dx.doi.org/10.1186/1743-422X-8-544.

Ma, J.-J., Zhao, F., Zhang, J., Zhou, J.-H., Ma, L.-N., Ding, Y.-Z., Chen, H.-T., Gu, Y.-X., Liu, Y.-S., 2013. Analysis of synonymous codon usage in dengue viruses. J. Anim. Vet. Adv. 12, 88–98.

Ma, Y.P., Liu, Z.X., Hao, L., Ma, J.Y., Liang, Z.L., Li, Y.G., Ke, H., 2015. Analysing codon usage bias of cyprinid herpesvirus 3 and adaptation of this virus to the hosts. J. Fish Dis. 38, 665–673. http://dx.doi.org/10.1111/jfd.12316.

Martins, V.E.P., Alencar, C.H., Kamimura, M.T., de Carvalho Araújo, F.M., de Simone, S.G., Dutra, R.F., Guedes, M.I.F., 2012. Occurrence of natural vertical transmission of dengue-2 and dengue-3 viruses in *Aedes aegypti* and *Aedes albopictus* in Fortaleza, Ceará, Brazil. PLoS One 7, 1–9. http://dx.doi.org/10.1371/journal.pone.0041386.

Metanat, M., Mood, B.S., Salehi, M., 2016. CCHF: A Reemeging Disease Crimean-Congo Hemorrhagic Fever: A Reemerging Disease. http://dx.doi.org/10.5812/archcid.20568.

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. http://dx.doi.org/10.1016/j.meegid.2013.01.001.

Morrison, A.C., Zielinski-Gutierrez, E., Scott, T.W., Rosenberg, R., 2008. Defining challenges and proposing solutions for control of the virus vector *Aedes aegypti*. PLoS Med. 5, 0362–0366. http://dx.doi.org/10.1371/journal.pmed.0050068.

Mueller, S., Papamichail, D., Coleman, J.R., 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. http://dx.doi.org/10.1128/JVI.00738-06.

Nasrullah, I., Butt, A.M., Tahir, S., Idrees, M., Tong, Y., 2015. Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. BMC Evol. Biol. 15, 174. http://dx.doi.org/10.1186/s12862-015-0456-4.

Novembre, J. a, 2000. Letter to the editor accounting for background nucleotide composition when measuring codon usage bias. Amino Acids 2, 1390–1394.

Papa, A., Ma, B., Kouidou, S., Tang, Q., Hang, C., 2002. Genetic characterization of the M RNA segment of Crimean Congo hemorrhagic fever virus strains. China 8.

Pena, D.J., Azevedo, V., Machado, C.R., 2009. Virus-Host Coevolution: Common Patterns of Nucleotide Motif Usage in Flaviviridae and Their Hosts. pp. 4. http://dx.doi.org/10.1371/journal.pone.0006282.

Peyrefitte, C., Marianneau, P., Tordo, N., Bouloy, M., 2015. Crimean-Congo haemorrhagic fever. 34, 391–401.

Puigbò, P., Bravo, I.G., Garcia-Vallve, S., 2008a. CAIcal: a combined set of tools to assess codon usage adaptation. Biol. Direct 3, 38. http://dx.doi.org/10.1186/1745-6150-3-38.

Puigbò, P., Bravo, I.G., Garcia-Vallvé, S., 2008b. E-CAI: a novel server to estimate an expected value of codon adaptation index (eCAI). BMC Bioinformatics 9, 65. http://dx.doi.org/10.1186/1471-2105-9-65.

Rabadan, R., Levine, A.J., Robins, H., 2006. Comparison of avian and human influenza A viruses reveals a mutational bias on the viral genomes. J. Virol. 80, 11887–11891. http://dx.doi.org/10.1128/JVI.01414-06.

Rahman, S.U., Mao, Y., Tao, S., 2017. Codon usage bias and evolutionary analyses of Zika virus genomes. Genes Genomics 0 (0). http://dx.doi.org/10.1007/s13258-017-0549-0.

Roychoudhury, S., Mukherjee, D., 2010. A detailed comparative analysis on the overall codon usage pattern in herpesviruses 148, 31–43. doi:https://doi.org/10.1016/j.virusres.2009.11.018.

Schubert, A.M., Putonti, C., 2010. Evolution of the sequence composition of Flaviviruses. Infect. Genet. Evol. 10, 129–136. http://dx.doi.org/10.1016/j.meegid.2009.11.004.

Serres-Giardi, L., Belkhir, K., David, J., Glemin, S., 2012. Patterns and evolution of nucleotide landscapes in seed plants. Plant Cell 24, 1379–1397. http://dx.doi.org/10.1105/tpc.111.093674.

Shackelton, L.A., Parrish, C.R., Holmes, E.C., 2006. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J. Mol. Evol. 62, 551–563. http://dx.doi.org/10.1007/s00239-005-0221-1.

Sharp, P.M., Li, W.H., 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for "rare" codons. Nucleic Acids Res. 14, 7737–7749. http://dx.doi.org/10.1093/nar/14.19.7737.

Sharp, P.M., Emery, L.R., Zeng, K., 2010. Forces that Influence the Evolution of Codon Bias. pp. 1203–1212. http://dx.doi.org/10.1098/rstb.2009.0305.

Shayan, S., Bokaean, M., Shahrivar, M.R., Chinikar, S., 2015. Crimean-Congo hemorrhagic fever. Lab. Med. 46, 180–189. http://dx.doi.org/10.1309/LMN1P2FRZ7BKZSCO.

Simmonds, P., 2012. SSE: a nucleotide and amino acid sequence analysis platform. BMC Res. Notes 550 (50), 50. http://dx.doi.org/10.1186/1756-0500-5-50.

Singh, N.K., Tyagi, A., Kaur, R., Verma, R., Gupta, P.K., 2016. Characterization of codon usage pattern and influencing factors in Japanese encephalitis virus. Virus Res. 221, 58–65. http://dx.doi.org/10.1016/j.virusres.2016.05.008.

Tao, P., Dai, L., Luo, M., Tang, F., Tien, P., Pan, Z., 2009. Analysis of synonymous codon usage in classical swine fever virus. Virus Genes 38, 104–112. http://dx.doi.org/10.1007/s11262-008-0296-z.

Wang, M., Liu, Y.-S., Zhou, J.-H., Chen, H.-T., Ma, L.-N., Ding, Y.-Z., Liu, W.-Q., Gu, Y.-X., Zhang, J., 2011. Analysis of codon usage in Newcastle disease virus. Virus Genes 42, 245–253. http://dx.doi.org/10.1007/s11262-011-0574-z.

Wang, H., Liu, S., Zhang, B., Wei, W., 2016. Analysis of Synonymous Codon Usage Bias of Zika Virus and Its Adaption to the Hosts. pp. 1–22. http://dx.doi.org/10.1371/journal.pone.0166260.

Weaver, S.C., 2006. Evolutionary Influences in Arboviral Disease, Quasispecies: Concept and Implications for Virology. Springer Berlin Heidelberg, Berlin, Heidelberg. http://dx.doi.org/10.1007/3-540-26397-7_10.

Wickham, H., 2016. ggplot2 — Elegant Graphics for Data Analysis. Springerhttp://dx.doi.org/10.18637/jss.v077.b02.

Wong, E.H.M., Smith, D.K., Rabadan, R., Peiris, M., Poon, L.L.M., 2010. Codon usage bias and the evolution of influenza A viruses. Codon usage biases of influenza virus. BMC Evol. Biol. 10, 253. http://dx.doi.org/10.1186/1471-2148-10-253.

World Health Organization, 2013. Crimean-Congo Haemorrhagic Fever. Fact Sheet. World Health Organization, Geneva Available from: http://www.who.int/mediacentre/factsheets/fs208/en/.

Wright, F., 1990. The "effective number of codons" used in a gene. Gene 87, 23–29. http://dx.doi.org/10.1016/0378-1119(90)90491-9.

Wu, Y., Zhao, D., Tao, J., 2015. Analysis of codon usage patterns in herbaceous Peony (*Paeonia lactiflora* Pall.) based on transcriptome data. Genes (Basel) 6, 1125–1139. http://dx.doi.org/10.3390/genes6041125.

Yurievna, N., Stanislavovna, I., Pavlovna, E., Borisovna, E., Sergeevich, A., 2016. International journal of infectious diseases possible sexual transmission of Crimean-Congo hemorrhagic fever. Int. J. Infect. Dis. 45, 109–111. http://dx.doi.org/10.1016/j.ijid.2016.02.1008.

Zhou, J.H., Gao, Z.L., Zhang, J., Chen, H.T., Pejsak, Z., Ma, L.N., Ding, Y.Z., Liu, Y.S., 2012. Comparative the codon usage between the three main viruses in pestivirus genus and their natural susceptible livestock. Virus Genes 44, 475–481. http://dx.doi.org/10.1007/s11262-012-0731-z.

Zhou, J. hua, Zhang, J., Sun, D. jie, Ma, Q., Chen, H. tai, Ma, L. na, Ding, Y. zhong, Liu, Y. sheng, 2013. The distribution of synonymous codon choice in the translation initiation region of dengue virus. PLoS One 8. http://dx.doi.org/10.1371/journal.pone.0077239.