Cellular and Molecular Biology
E-ISSN : 1165-158X / P-ISSN : 0145-5680
www.cellmolbiol.org

Original Research

Analysis of codon usage bias in the VP1 gene of the human norovirus GII.2 genotype

Haoning Wang[1], Guanglin Cui[2], Xiaolong Wang[1, 3#*], Minghai Zhang[1#*], Jimin Tan[4#*], Xingguang Li[5#*]

[1] College of Wildlife Resource, Northeast Forestry University, 26 Hexing Road, Xiangfang Distract, Harbin 150040, Heilongjiang, China
[2] Key Clinical Laboratory of Henan Province, Department of Clinical Laboratory, First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, Henan, China
[3] Center of Conservation Medicine and Ecological Safety, Northeast Forestry University, 6 Haping Road, Xiangfang Distract, Harbin 150040, Heilongjiang, China
[4] Jiangsu Engineering Laboratory of Animal Immunology, Institute of Immunology and College of Veterinary Medicine, Nanjing Agricultural University, Nanjing 210095, China
[5] Hubei Engineering Research Center of Viral Vector, Wuhan University of Bioengineering, Wuhan, 430415, China

Correspondence to: yttuhh@yeah.net

Received October 14, 2018; **Accepted** December 1, 2018; **Published** December 31, 2018

[#]These authors contributed equally to this work.

Doi: http://dx.doi.org/10.14715/cmb/2017.64.15.18

Abstract: To investigate the codon usage patterns of all available VP1 gene sequences of the GII.2 genotype, to determine the factors that affect these patterns, and to provide comprehensive details of the characteristics and evolution of the gene. Complete 519 sequences of VP1 gene of the HuNoV GII.2 genotype with known sampling dates and geographic locations from 1971 - 2017 were retrieved from the GenBank nucleotide database of the National Center for Biotechnology Information (NCBI) and analyzed. The percentage composition of T, C, A, and G nucleotides were 24.80 ± 0.30, 26.61 ± 0.31, 25.84 ± 0.13, and 22.75 ± 0.17 %, respectively, with C and A relatively more abundant than T and G, and C the most abundant ($p < 0.0001$). The values of T3s (34.10 ± 0.90 %) and C3s (33.54 ± 0.90 %) were significantly higher than those of A3s (29.98 ± 0.43 %) and G3s (24.13 ± 0.51 %) ($p < 0.0001$). While T3s was highest among the four nucleotides, G3s was the lowest. Among the 18 most frequently employed synonymous codons, six optional codons ended with T, five ended with C, five ended with A and two ended with G. Codons ending with T were the most frequently used. The ENC ranged from 51.90 to 54.25 (mean = 52.38 ± 0.43) among the 519 VP1 gene sequences. There were significant correlations between ENC and C % and G % ($p < 0.01$). Codons containing CpG (1 and 2 or 2 and 3 codon positions) showed the lowest frequencies, while 30, 29, and 2 codons were above, below and on the mean line, respectively. The first four principal components accounted for 69.11 % of the total variation, with the first, second, third, and fourth principal axes contributing 37.90, 14.83, 9.61, and 6.77 %, respectively. The strains were not clustered by country of isolation or year of sampling. Gravy were significantly correlated with T3s, C3s, G3s, GC3s, and ENC ($p < 0.01$). Mutation pressure and natural selection contributed to the codon usage bias of the VP1 gene of the HuNoV GII.2 genotype. There was a correlation between GC12s and GC3s ($R^2 = 0.032$; $p < 0.0001$). The relative neutrality was 3.20 %, while natural selection was 96.80 %. The VP1 gene exhibits low codon usage bias which is affected primarily by natural selection, followed by mutation pressure and translational selection.

Key words: Human norovirus; GII.2 genotype; Codon usage bias; Natural selection; Mutation pressure.

Introduction

Human norovirus (HuNoV) infections occur frequently and are a major cause of gastroenteritis (1 - 3). The challenge posed by these infections to global health is aggravated by factors such as immune-compromise and globalization of the food industry (4, 5). The high diversity of genetic and antigenic noroviruses complicate the development of vaccines and therapies required to combat these challenges (6, 7). Human norovirus belongs to the *Caliciviridae* family, and possesses a single-stranded positive-sense RNA genome (7.5 kb) with three open reading frames (ORFs) that encode different proteins (8). For instance, ORF1 encodes a large polyprotein which can be cleaved into smaller proteins such as p22, p28, P-loop NTPase, VPg, 3C-like protease, and RNA-dependent RNA polymerase; while ORF2 and ORF3 encode major and minor structural proteins (VP1 and VP2, respectively). The VP1 has a shell (S) and protruding (P) domains. The P domain contains elements of antigenicity due to its ability to bind to histo-blood group antigens (HBGAs) which serve as co-receptors on host cells (9, 10). Repeated recombination within the ORF1-ORF2 overlap region and antigenic variation in the P domain are responsible for the appearance of novel viral strains that can evade host immunity (3, 11). Codons that encode the same amino acids are referred to as synonymous codons, and their use is a non-random process, with some codons being utilized more often than others. This phenomenon is called “codon usage bias” and it is found in prokaryotes, eukaryotes, and viruses (12). Codon usage is primarily influenced by natural selection and mutational bias, and usage between viruses and hosts can affect overall virus survival, ability to evade host immunity, and viral evolution (13, 14). Thus, an understanding of codon usage by viruses can

provide valuable information about their evolution and expand our understanding of the regulation of viral gene expression based on codon adoption (15). This can in turn aid the design of vaccines aimed at bringing about lasting immunity (16). Since different strains of HuNoV cause diseases with a wide range of clinical signs, it is important to characterize their genetic variation, evolution, and codon usage patterns so as to understand how these viral strains cause diseases. The aim of this study was to investigate the codon usage patterns of all available VP1 gene sequences of the GII.2 genotype, determine the factors that affect these patterns, and provide comprehensive details of the characteristics and evolution of the gene.

Materials and Methods

Sequence dataset

Complete sequences of VP1 gene of the HuNoV GII.2 genotype with known sampling dates and geographic locations from 1971 - 2017 were retrieved from the GenBank nucleotide database of the National Center for Biotechnology Information (NCBI) and used for this study. The final dataset included 519 VP1 gene sequences with sampling locations from Australia (n = 3), China (n = 371), Hong Kong (n = 22), Japan (n = 106), Malaysia (n = 2), Taiwan (n = 9), and USA (n = 7) (Supplementary Table S1). All dataset sequences were aligned with Multiple Alignment using Fast Fourier Transform (MAFFT) (7.222), and further manually adjusted using BioEdit (7.2.5). Multiple sequence alignments were screened for recombinant viral sequences with no recombinant sequences using RDP (4.36).

Analysis of nucleotide composition

The percentage composition of each nucleotide and frequency of occurrence were calculated using BioEdit (version 7.2.5). The nucleotide composition of the third synonymous codon position (T3s, C3s, A3s, and G3s) in all the sequences was calculated using CodonW (http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::CodonW). In addition, the nucleotide compositions of G + C at the first, second, third, and all codon positions (GC1s, GC2s, GC3s, and GCs) for all sequences were calculated using CAT (1.0) with 10,000 bootstrap replicates (Supplementary Table S2)

Determination of relative synonymous codon usage (RSCU)

The RSCU index was calculated using the formula:

$$RSCU = \frac{g_{ij}}{\sum_j^{n_i} g_{ij}} n_i$$

where g_{ij} is the observed number of the i^{th} codon for the j^{th} amino acid with n_i synonymous codons. Values of RSCU equivalent to 1.0, > 1.0, or < 1.0, represent no bias, positive codon usage bias, and negative codon usage bias, respectively.

Determination of effective number of codons (ENC)

This was calculated using the formula:

$$ENC = 2 + \frac{9}{\bar{F}_2} + \frac{1}{\bar{F}_3} + \frac{5}{\bar{F}_4} + \frac{3}{\bar{F}_6}$$

where $\bar{F}_i$ (i = 2, 3, 4, 6) represents the mean of values for the i-fold degenerate amino acids. The equation below was used to calculate the $\bar{F}_i$ values:

$$\bar{F}_i = \frac{n\sum_{j=1}^{i}(\frac{n_j}{n})^2 - 1}{n-1}$$

where n is the total number of occurrences of the codon for that amino acid, and n_i is the total number of occurrences of the j^{th} codon for that amino acid represented by n_j.

To determine the relationship between the observed GC3s and ENC, the expected ENC for different GC3s was calculated using the formula:

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

where "s" represents the possible value of GC3s. The observed and expected ENC were compared to determine the influence of nucleotide compositional constraint on shaping synonymous codon usage bias, and to clarify the effect of natural selection pressure. Plots of ENC were constructed to determine the factors that influence codon usage bias.

Measurement of relative abundances of dinucleotides and frequencies of codons

The relative abundances of the 16 dinucleotides in the VP1 gene were calculated using CodonW (1.4.2), as shown in the formula below (17):

$$P_{xy} = \frac{f_{xy}}{f_x f_y}$$

where f_x and f_y represent the frequencies of nucleotides X and Y, respectively, and f_{xy} and $f_y f_x$ represent the observed and expected frequencies of the dinucleotides, respectively. Values of P_{xy} < 0.78 or > 1.23 were considered as under- or over-represented, respectively (18). The frequencies of the 64 codons of the VP1 gene were also calculated using CodonW.

Analysis of neutrality plot

This was carried out according to the method described by Fuglsang (19). In the neutrality plots, GC12s were plotted against GC3s, with each dot representing an independent VP1 gene of the HuNoV GII.2 genotype. Based on linear regression analysis, a slope parallel to the abscissa axis indicates that only natural selection constrains the codon bias, whereas a line near the vertical axis indicates complete mutation pressure, and a line that coincides with the diagonal indicates complete neutrality.

Analyses of general average hydrophilicity (Gravy) and aromaticity (Aroma)

To determine if natural selection shaped codon usage bias, Gravy and Aroma values were calculated using CodonW (http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::CodonW), with higher Gravy or Aroma values indicating more hydrophobic or aromatic amino acid products, respectively (Supplementary Table S2) .

Principal component analysis (PCA)

This was performed on the VP1 gene of the GII.2

genotype. Major trends in the codon usage patterns of the VP1 gene of the GII.2 genotype were subjected to principal component analysis (PCA) based on country of isolation and year of sampling, using GraphPad Prism v.7.04.

Statistical analysis

Measurement data are expressed as mean ± SD. All statistical analyses were performed using GraphPad Prism (7.04). Correlation and significance analyses were performed using Spearman's rank correlation and Chi-squared tests, respectively. Values of $p < 0.0001$, 0.01, or 0.05 were considered statistically significant.

Results

Nucleotide composition of the VP1 gene of the HuNoV GII.2 genotype

The percentage composition of T, C, A, and G were 24.80 ± 0.30, 26.61 ± 0.31, 25.84 ± 0.13, and 22.75 ± 0.17 %, respectively. The C and A were more relatively abundant than T and G, with C being the most abundant nucleotide ($p < 0.0001$). The values of T3s (34.10 ± 0.90 %) and C3s (33.54 ± 0.90 %) were significantly higher than those of A3s (29.98 ± 0.43 %) and G3s (24.13 ± 0.51 %) ($p < 0.0001$). While T3s was highest among the four nucleotides, G3s was the lowest (Table 1).

Outcomes of RSCU and ENC analyses

Among the 18 most frequently employed synonymous codons, six optional codons ended with T (TGT for Cys, GGT for Gly, ATT for Ile, AAT for Asn, TCT for Ser, and TAT for Tyr); five ended with C (GCC for Ala, GAC for Asp, TTC for Phe, CAC for His, and CTC for Leu); five ended with A (GAA for Glu, AAA for Lys, CCA for Pro, AGA for Arg, and ACA for Thr), while two ended with G (CAG for Gln and GTG for Val). Codons ending with T were the most frequently used (Table 2).

The ENC ranged from 51.90 to 54.25 (mean = 52.38 ± 0.43) among the 519 VP1 gene sequences.

Role of mutational bias in shaping codon usage of the VP1 gene

On plotting ENC against GC3s according to the geographical distribution and year of sampling of the VP1 gene of the GII.2 genotype, all the points were located below the expected curve (Figure 1). Most strains from different countries were clustered together (Figure 1A), as were most strains sampled from 1971 to 2017 (Figure 1B). The percentage compositions of nucleotides (T, C, A, G, and GC), and codon content (T3s, C3s, A3s, G3s, and GC3s) were mix-correlated ($p < 0.01$), except for A3s and T %, G3s and T %, GC3s and A %, and GC3s and G %, indicating that most nucleotide compositions were related to the third codon position (Table 2). Furthermore, there were significant correlations between ENC and C and G ($p < 0.01$). The frequency distributions of all 61 codons were not homogeneous (Figure S1). Codons containing CpG (1 and 2 or 2 and 3 codon positions) showed the lowest frequencies, while 30, 29, and 2 codons were above, below and on the mean line, respectively (Figure S1). To determine the possible effect of the dinucleotides on codon usage, the relative abundances of the 16 dinucleotides of the VP1 gene of the GII.2 genotype were calculated (Figure S2a). Results showed that TpG (1.469 ± 0.030), CpC (1.321 ± 0.021), CpA (1.292 ± 0.024), and GpG (1.279 ± 0.025) exhibited different degrees of over-representation, while TpA (0.634 ± 0.020), CpG (0.349 ± 0.018), and GpC (0.753 ± 0.020) showed different degrees of under-representation. It is worth noting that CpG showed the lowest abundance among all the dinucleotides, consistent with the trinucleotide analysis which revealed that the codons contained the lowest frequency of CpG (Figure S1). The RSCU values of codons containing CpG (CGA, CGU, CGC, TCG, ACG, CGG, GCG, and CCG) were the lowest among the synonymous codons. The strong variation in the frequencies of the 20 amino acids and the relatively large fluctuation in the abundances of the 16 dinucleotides indicated that translational selection also shaped codon usage patterns of the VP1 gene

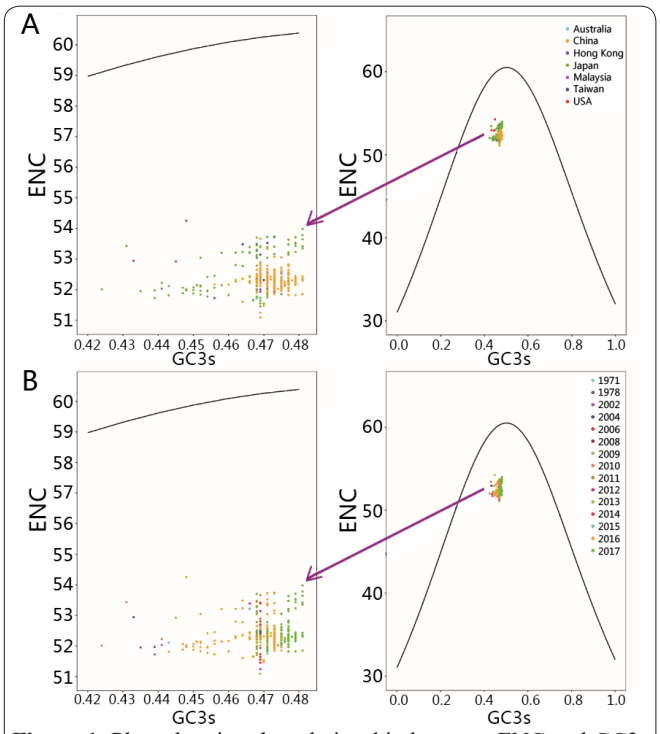

Figure 1. Plots showing the relationship between ENC and GC3s according to the country of isolation (A), and year of sampling (B). Different countries of isolation and years of sampling are represented in different colors.

Table 1. Correlation analyses of nucleotide composition, Gravy, Aroma, Axis 1, Axis 2, third position nucleotide, and ENC values.

	T %	C %	A %	G %	GC	Gravy	Aroma	Axis 1	Axis 2
T3s	0.911**	-0.917**	-0.300**	0.299**	-0.856**	0.192**	0.035	-0.206**	0.018
C3s	-0.887**	0.924**	0.290**	-0.348**	0.843**	-0.179**	-0.055	0.189**	-0.068
A3s	-0.027	0.150**	0.692**	-0.739**	-0.318**	0.084	-0.067	-0.222**	-0.163**
G3s	-0.076	-0.147**	-0.491**	0.764**	0.310**	-0.181**	0.101*	0.438**	0.146**
GC3s	-0.881**	0.830**	0.029	0.007	0.959**	-0.239**	-0.014	0.324**	0.029
ENC	-0.053	0.236**	0.037	-0.363**	0.023	0.314**	0.002	-0.382**	-0.614**

Table 2. Synonymous codon usage patterns in the VP1 gene.

AA	Codon	RSCU/Number	AA	Codon	RSCU/Number
A (Ala)	GCA	0.844/4702	P (Pro)	***CCA***	***1.514/10011***
A	***GCC***	***1.997/11126***	P	CCC	0.745/4929
A	GCG	0.191/1064	P	CCG	0.290/1919
A	GCT	0.969/5398	P	CCT	1.451/9597
C (Cys)	TGC	0.958/1242	Q (Gln)	CAA	1.183/7688
C	***TGT***	***1.042/1351***	Q	***CAG***	***1.634/5307***
D (Asp)	***GAC***	***1.011/6849***	R (Arg)	***AGA***	***2.962/6193***
D	GAT	0.989/6694	R	AGG	2.044/4273
E (Glu)	***GAA***	***1.150/6833***	R	CGA	0.161/336
E	GAG	0.850/5053	R	CGC	0.247/516
F (Phe)	***TTC***	***1.224/8896***	R	CGG	0.245/512
F	TTT	0.776/5642	R	CGT	0.342/715
G (Gly)	GGA	0.778/4029	S (Ser)	AGC	0.624/1575
G	GGC	0.565/2924	S	AGT	1.026/2589
G	GGG	1.001/5183	S	TCA	0.830/2094
G	***GGT***	***1.657/8582***	S	TCC	1.495/3773
H (His)	***CAC***	***1.090/3666***	S	TCG	0.410/1034
H	CAT	0.910/3061	S	***TCT***	***1.615/4074***
I (Ile)	ATA	0.610/2144	T (Thr)	***ACA***	***1.902/8684***
I	ATC	1.143/4017	T	ACC	1.097/5007
I	***ATT***	***1.247/4385***	T	ACG	0.023/103
K (Lys)	***AAA***	***1.050/3220***	T	ACT	0.979/4469
K	AAG	0.950/2911	V (Val)	GTA	0.116/721
L (Leu)	CTA	0.752/3058	V	GTC	1.257/7787
L	***CTC***	***1.705/6934***	V	***GTG***	***1.883/11664***
L	CTG	0.929/3779	V	GTT	0.743/4604
L	CTT	1.019/4144	W (Trp)	TGG	1.000/2595
L	TTA	0.512/2084	Y (Tyr)	TAC	0.900/3969
L	TTG	1.083/4405	Y	***TAT***	***1.100/4849***
M (Met)	ATG	1.000/6741			
N (Asn)	AAC	0.775/7593			
N	***AAT***	***1.225/12002***			

of the HuNoV GII.2 genotype (Figure S1 and S2a).

Outcomes of PCA

The first four principal components accounted for 69.11 % of the total variation, with the first, second, third, and fourth principal axes contributing 37.90, 14.83, 9.61, and 6.77 %, respectively (Figure S2b). The strains were not clustered by country of isolation (Figure 2A), or year of sampling (Figure 2B).

Role of natural selection in shaping codon usage patterns of the VP1 gene of the HuNoV GII.2 genotype

Gravy were significantly correlated with T3, C3, G3, GC3, and ENC ($p < 0.01$) (Table 2). Mutation pressure and natural selection contributed to the codon usage bias of the VP1 gene of the HuNoV GII.2 genotype. There was correlation between GC12s and GC3s (0.032; $p < 0.0001$). The relative neutrality was 3.20 %, while natural selection was 96.80 % (Figure S3).

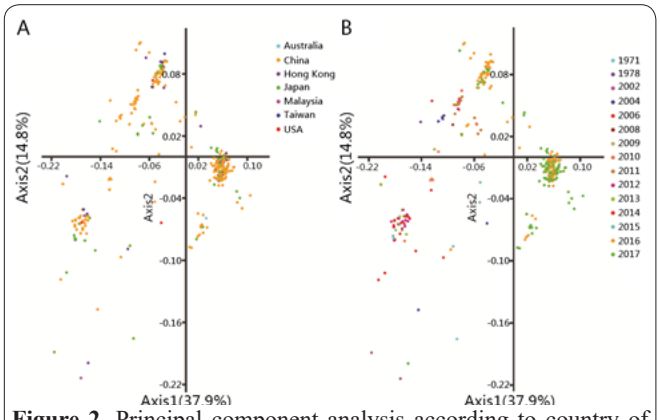

Figure 2. Principal component analysis according to country of isolation (A), and year of sampling (B) for each sequence of the VP1 gene. Different countries of isolation and years of sampling are represented in different colors.

Discussion

Since 1990, GII.4 has been the predominant HuNoV genotype in humans and this dominance is attributed to the continuous replacement of antigenically distinct intra-genotype variants which can escape herd immunity (20-22). No other genotypes appear to possess this advantage as their VP1 amino acid sequences remain almost invariable or static despite decades of circulation

in the human population (23, 24). In the winters of 2014 and 2015, the GII.17 genotype unexpectedly replaced GII.4, emerging as a major cause of acute gastroenteritis in several Asian countries (25). During the winter of 2016 to 2017, the GII.2 genotype became dominant in Germany, China, Japan, Taiwan, and Hong Kong in HuNoV-associated acute gastroenteritis epidemics (26-28).

The VP1 gene is the most widely studied of the HuNoV genome and it consists of S and P domains. The P domain exhibits significant determinants of antigenicity by binding to HBGAs which function as attachment factors or co-receptors on host cells (9, 29). This makes it an essential component of pathogenicity and the primary target of antibodies (30). However, by probing information contained in sampled viral sequences, few studies have comprehensively assessed the codon usage patterns of the VP1 gene of the HuNoV GII.2 genotype.

In the present study, C and A were more relatively abundant than T and G, with C being the most abundant nucleotide. The codon compositions of the synonymous codons at the third position showed that the values of T3s were higher than those of A3s and G3s. The T3s value was the highest among the four nucleotides, which was inconsistent with the nucleotide content of the VP1 gene of the GII.2 genotype, while the G3s was the lowest, and was inconsistent. Among the 18 most frequently employed synonymous codons, six optional codons ended with T, five ended with C, five ended with A, and two ended with G. Codons ending with T were the most frequently used. These results appear to suggest that codon usage may be significantly influenced by compositional constraints.

In this study, the ENC was greater than 45, a possible indication of low fluctuation and low codon usage bias among the 519 VP1 gene sequences. The plot of ENC against GC3s according to the geographical distribution and year of sampling of the VP1 gene of the GII.2 genotype suggested that low codon usage bias may be affected by nucleotide composition. These results also suggest that codon usage patterns of the VP1 gene may be strongly affected by natural selection pressure. In addition, nucleotide composition and codon content were mix-correlated, except for A3s and T %, G3s and T %, GC3s and A %, and GC3s and G %. These results suggest that most nucleotide compositions were probably related to the third codon position. There were significant correlations between ENC and C % and G %, respectively. These results have shown that low codon usage bias patterns of the VP1 gene of the HuNoV GII.2 genotype may be affected by mutational bias.

In this study, analyses of frequencies of the 61 codons encoding all 20 amino acids and the relative abundances of the 16 dinucleotides showed that translational selection may also play an important role in shaping codon usage patterns. Analysis of the relationships between nucleotide composition and codon content at the third base positions suggested that mutation pressure probably shaped codon usage of the VP1 gene of the GII.2 genotype. Analysis of Gravy, Aroma, and codon content showed correlations between Gravy and T3s, G3s, and GC3s, further supporting the claim that natural selection may contribute significantly to the codon usage of the VP1 gene. The RSCU of codons containing CpG were the lowest among the synonymous codons. These results suggest that translational selection may also shape codon usage patterns of the VP1 gene. Results of PCA showed that the first and second axes primarily contributed to the variation in the RSCU of synonymous codons and strains were not clustered by country of isolation or year of sampling. This result appears to suggest that mutation pressure and natural selection may contribute to codon usage patterns of the VP1 gene of the HuNoV GII.2 genotype. It was found that natural selection was the major force driving codon usage of the VP1 gene of the HuNoV GII.2 genotype. Hence, except for some properties of the sequences themselves, the codon usage patterns and sequence characteristics of the VP1 gene of the HuNoV GII.2 genotype were restricted by the evolutionary process. The plot of GC12s against GC3s revealed significant correlation between them, suggesting that natural selection may play a greater role in shaping codon usage bias of VP1 gene of the HuNoV GII.2 genotype than mutation pressure. The results obtained in this study suggest that monitoring updated sequences of this novel emerging virus will help to better understand the evolution of viral sequences and disease correlated with T3s, C3s, G3s, GC3s, and ENC. This suggests that natural selection may markedly influence codon usage patterns of the VP1 gene and probably play a more important role than mutation pressure in shaping codon usage.

These results indicate that VP1 gene exhibits low codon usage bias which is affected primarily by natural selection, followed by mutation pressure, and translational selection. The primary element affecting codon usage patterns of the VP1 gene is natural selection. This is the first study to analyze codon usage bias of the VP1 gene of the HuNoV GII.2 genotype and describe the forces that drive VP1 gene evolution. Future epidemiological surveys and sequence analyses are required to investigate the factors that drive the evolution of the HuNoV GII.2 genotype.

Acknowledgements
The authors wish to specially thank Dr. Christine Watts from Australia for her help in the preparation of the manuscript.

Funding
This work was supported with grants from the Fundamental Research Funds for the Central Universities (Code: 2572017PZ11).

Conflict of interest
The authors declare that no conflict of interest is associated with this work.

Author contributions
Xiaolong Wang, Minghai Zhang, Jimin Tan and Xingguang Li designed the research. Jimin Tan and Xingguang Li analysed the data and drafted the manuscript. Haoning Wang, Guanglin Cui, Xiaolong Wang, Minghai Zhang, Jimin Tan and Xingguang Li interpreted data and provided critical comments. All authors reviewed and approved the final manuscript.

References

1. Patel MM, Widdowson MA, Glass RI, Akazawa K, Vinje J, Parashar UD. Systematic literature review of role of noroviruses in sporadic gastroenteritis. Emerg Infect Dis 2008; 14: 1224-1231.
2. Atmar RL, Opekun AR, Gilger MA, Estes MK, Crawford SE, Neill FH, et al. Norwalk virus shedding after experimental human infection. Emerg Infect Dis 2008; 14: 1553-1557.
3. Ruvoen-Clouet N, Belliot G, Le Pendu J. Noroviruses and histo-blood groups: the impact of common host genetic polymorphisms on virus transmission and evolution. Rev Med Virol 2013; 23: 355-366.
4. Sakon N, Sadamasu K, Shinkai T, Hamajima Y, Yoshitomi H, Matsushima Y, et al. Foodborne Outbreaks Caused by Human Norovirus GII.P17-GII.17-Contaminated Nori, Japan, 2017. Emerg Infect Dis 2018; 24: 920-923.
5. Swartling L, Ljungman P, Remberger M, Sundin M, Tiveljung A, Mattsson J, et al. Norovirus causing severe gastrointestinal disease following allogeneic hematopoietic stem cell transplantation: A retrospective analysis. Transpl Infect Dis 2018; 20: 12847.
6. Dunkin N, Weng SC, Coulter CG, Jacangelo JG, Schwab KJ. Impacts of virus processing on human norovirus GI and GII persistence during disinfection of municipal secondary wastewater effluent. Water Res 2018; 134: 1-12.
7. Boonchan M, Guntapong R, Sripirom N, Ruchusatsawat K, Singchai P, Rungnobhakhun P, et al. The dynamics of norovirus genotypes and genetic analysis of a novel recombinant GII.P12-GII.3 among infants and children in Bangkok, Thailand between 2014 and 2016. Infect Genet Evol 2018; 60: 133-139.
8. Thongprachum A, Okitsu S, Khamrin P, Maneekarn N, Hayakawa S, Ushijima H. Emergence of norovirus GII.2 and its novel recombination during the gastroenteritis outbreak in Japanese children in mid-2016. Infection Genetics and Evolution 51:86-88.
9. Hardy ME. 2005. Norovirus protein structure and function. Fems Microbiol Letter 2017; 253: 1-8.
10. Donaldson EF, Lindesmith LC, Lobue AD, Baric RS. Viral shape-shifting: norovirus evasion of the human immune system. Nat Rev Microbiol 2010; 8: 231-241.
11. Siqueira JAM, Sousa EC, Linhares AD, Gabbay YB. Molecular analysis of norovirus in specimens from children enrolled in a 1982-1986 study in Belem, Brazil: A community-based longitudinal study. J Med Virol 2017; 89: 1539-1549.
12. Comeron JM, Aguade M. An evaluation of measures of synonymous codon usage bias. J Mol Evol 1998; 47: 268-274.
13. Sorensen MA, Kurland CG, Pedersen S. Codon Usage Determines Translation Rate in Escherichia-Coli. J Mol Biol 1989; 207: 365-377.
14. Eyrewalker AC. An Analysis of Codon Usage in Mammals - Selection or Mutation Bias. J Mol Evol 1991; 33: 442-449.
15. Schachtel GA, Bucher P, Mocarski ES, Blaisdell BE, Karlin S. Evidence for Selective Evolution in Codon Usage in Conserved Amino-Acid Segments of Human Alphaherpesvirus Proteins. J Mol Evol 1991; 33: 483-494.
16. Seetharaman J, Srinivasan R. Analysis of Codon Usage - Positional Preference in Various Organisms. Indian J Biochem Biophys 1995; 32: 156-160.
17. Karlin S, Burge C. Dinucleotide Relative Abundance Extremes - a Genomic Signature. Trends Genet 1995; 11: 283-290.
18. Butt AM, Nasrullah I, Qamar R, Tong YG. Evolution of codon usage in Zika virus genomes is host and vector specific. Emerg Microbes Infect 2016; 5: 107.
19. Fuglsang A. Impact of bias discrepancy and amino acid usage on estimates of the effective number of codons used in a gene, and a test for selection on codon usage. Gene 2008; 410: 82-88.
20. Fu JG, Ai J, Qi X, Zhang J, Tang FY, Zhu YF. Emergence of two novel norovirus genotype II.4 variants associated with viral gastroenteritis in China. J Med Virol 2014; 86: 1226-1234.
21. Godoy P, Artigues A, Bartolome R, Dominguez A, Plasencia A. Norovirus gastroenteritis outbreak by person to person transmision in a nursing home. Med Clin 2006; 127: 538-541.
22. Pothier P, Kaiser L. Norovirus disease today. Clin Microbiol Infect 2014; 20: 16.
23. Koo HS, Lee MO, Ku PT, Hwang SJ, Park DJ, Baik HS. Molecular epidemiology of norovirus in asymptomatic food handlers in Busan, Korea, and emergence of genotype GII.17. J Microbiol 2016; 54: 686-694.
24. Kobayashi M, Matsushima Y, Motoya T, Sakon N, Shigemoto N, Okamoto-Nakagawa R, et al. Molecular evolution of the capsid gene in human norovirus genogroup II. Sci Rep 2016; 6: 29400.
25. Lee CC, Feng Y, Chen SY, Tsai CN, Lai MW, Chiu CH. Emerging Norovirus GII. 17 in Taiwan. Clin Infect Dis 2015; 61: 1762.
26. Li JS, Qin M, Dong XG, Yang JY, Yang XX, Wei XX, et al. Norovirus outbreaks in Fengtai District, Beijing, China, 2014. Arch Virol 2016; 161: 2855-2858.
27. Polo TD, Peiro JR, Mendes LCN, Ludwig LF, de Oliveira EF, Bucardo F, et al. Human norovirus infection in Latin America. J Clin Virol 2016; 78: 111-119.
28. Rocha-Pereira J, Van Dycke J, Neyts J. Norovirus genetic diversity and evolution: implications for antiviral therapy. Curr Opin Virol 2016; 20: 92-98.
29. Fretz R. How is Norovirus-assosiated gastroenteritis important? Wien Klin Wochenschr 2005; 117: 785-788.
30. Larsson MM, Rydell GEP, Rodriguez-Diaz J, Akerlind B, Hutson AM, Estes MK, et al. Antibody prevalence and titer to norovirus (genogroup II) correlate with secretor (FUT2) but not with ABO phenotype or Lewis (FUT3) genotype. J Infect Dis 2006; 194: 1422-1427.