Journal of Animal and Veterinary Advances 11 (18): 3281-3288, 2012
ISSN: 1680-5593

Analysis of Codon Usage and Nucleotide Compositional Traits of Morbilliviruses

Feng-Chao Yan, Yong-Xi Dou and Lei Chen
State Key Laboratory of Veterinary Etiological Biology,
Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute,
Chinese Academy of Agricultural Sciences, 730046 Lanzhou, P.R. China

Abstract: In this study, the Relative Synonymous Codon Usage (RSCU) values, Effective Number of Codon (ENC) values, nucleotide contents and dinucleotide were used to consider codon usage pattern of each structure protein gene and genome among the members of *Morbillivirus* genus. The results showed that the overall extent of codon usage bias in Morbilliviruses is low (mean ENC = 54.84>35). The good correlation between the (C+G) 12% and (G+C) 3% suggested that the mutational pressure rather than natural selection is the main factor that determines the codon usage bias and base component in Morbilliviruses. It is observed that synonymous codon usage pattern in Morbilliviruses can be divided into two groups. Under-represented CpG is a characteristic of all the members in Morbilliviruses. It may be a strategy to adapt to the host. These analyses not only provided an insight into the variation of codon usage pattern among the genomes of Morbilliviruses but also were helpful in understanding the processes of the evolution of Morbilliviruses.

Key words: Morbillivirus, nucleotide composition, codon usage bias, host, protien

INTRODUCTION

Morbilliviruses infections are acute and highly contagious viral diseases causing high morbidity rates and sometimes high mortality rates. They are enveloped, non-segmented, negative-stranded RNA viruses which belong to the subfamily Paramyxovirinae in the family Paramyxoviridae. According to the report of International Committee Taxonomy Viruses (ICTV), six distinct species of Morbilliviruses have been recognized namely: Measles Virus (MV), Rinderpest Virus (RPV), Peste-des-Petits Ruminants Virus (PPRV), Canine Distemper Virus (CDV), Phocine Distemper Virus (PDV) and Cetacean Morbillivirus (CeMV). Members of Morbillivirus have a wide range of hosts that infect ruminants, dogs, marine mammals and humans.

Structural proteins of Morbilliviruses consist of Nucleocapsid (N) protein, Phosphoprotein (P) Fusion (F) protein, Hemagglutinin (H) protein, Matrix (M) protein and polymerase (L) protein. Each of the six is encoded by a separate gene (Diallo, 1990). The major protein in the virion is the Nucleocapsid protein (N) which encapsidates the negative-strand RNA genome in a Ribonucleoprotein Complex (RNP). The RNP also contains the P and L proteins. The non-structural proteins V and C are derived from the P gene by respectively insertion of an extra G residue in the mRNA (editing) and by translation of an Overlapping Reading Frame (ORF) in the same mRNA (Bellini *et al*., 1985; Cattaneo *et al*., 1989). All genes of the genome are arranged in the order 3'-N-P/V/C-M-F-H-L-5'. It was noticed that synonymous codons were used unequally in different genomes even in different genes of the same genome when molecular sequence data started to be accumulated nearly 20 years ago (Supek and Vlahovicek, 2005). Synonymous codon usage bias is an important evolutionary phenomenon and exists in a wide range of biological systems from prokaryotes to eukaryotes and to viruses (Archetti, 2004; Liu *et al*., 2010). Codon usage analysis had been applied to prokaryote and eukaryote such as *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and human beings (Bulmer, 1988; Kanaya *et al*., 2001; Karlin and Mrazek, 1996; Sharp *et al*., 1988). The codon usage bias is attributable to the equilibrium between natural selection and mutation pressure (Shackelton *et al*., 2006).

In recent studies of viral codon usage, researchers showed that mutation bias may be a more important factor than natural selection in determining codon usage bias of some viruses such as Picornaviridae, Pestivirus, plant

Corresponding Author: Yong-Xi Dou, State Key Laboratory of Veterinary Etiological Biology,
Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute,
Chinese Academy of Agricultural Sciences, 730046 Lanzhou, P.R. China

viruses and vertebrate DNA viruses (Fu, 2010; Roychoudhury *et al*., 2010; Shackelton *et al*., 2006; Tao *et al*., 2009; Zhong *et al*., 2007). Moreover, a number of evidences proved that codons using abundant tRNA are selectively favored, especially in highly expressed genes (Fu, 2010). Recently, it was also suggested that codon usage is correlated to gene function and used to evaluate the extent of bias toward codons that were known to be preferred in highly expressed genes (Epstein *et al*., 2000; Gu *et al*., 2004b). The information of codon usage include the Codon Adaptation Index (CAI), the G+C content at different position, the Relative Synonymous Codon Usage (RSCU) and the Effective Number of Codon (ENC). The parameters can reveal much about the molecular evolution or individual genes of the viruses and may also help to understand the regulation of viral gene expression (Carbone *et al*., 2003; Gu *et al*., 2004a; Jenkins and Holmes, 2003).

The codon usage pattern of the PPRV revealed that the codons ended with A or G are most preferred especially for *N* gene. The results showed that mutation bias plays a decisive role in evolution process of PPRV codon usage patterns. The researchers also, showed that the codon usage of all viruses in Morbillivirus have intraspecies differences (Liu *et al*., 2011). In the present study, it was the first time to study the codon usage pattern of Morbilliviruses as a whole and tried to clarify the molecular evolution of Morbilliviruses on the basis of their codon usage patterns.

MATERIALS AND METHODS

Sequence data: Completely sequenced 5 Morbilliviruses genomes and Coding Sequence (CDS) available from NCBI-GenBank were downloaded and analyzed in this study (Table 1). A total of 30 annotated ORFs from all these sequenced Morbilliviruses were used for further study.

Composition analysis of coding region of Morbilliviruses: To understand the influence of nucleotide composition on the genes and their codon usages of Morbilliviruses, A, T, G, C as well as A3, T3, G3, C3, GC3 and GC3s content of each gene were computed.

The GC content is a measure of the nucleotide composition of the whole coding region of a gene. GC3s is a measure of the nucleotide composition of only the third position of the codon (from 5-3 end of the gene) in addition to methionine, tryptophan and stop codon. It reflects the strength of directional mutation pressure and the codon preference. The values of the nucleotide composition were calculated by the online program CAIcal SERVER, respectively.

The dinucleotide values are an important indicator of the nature of compositional bias of the genes and they affect the codon usage bias. In order to understand the patterns of dinucleotide usage, the dinucleotide frequency was assessed for different viruses in Morbillivirus. Dinucleotide frequency ≥ 1.23 or ≤ 0.78 are considered as significantly over or under represented (Karlin *et al*., 2002).

The actual and predicted values of the Effective Number of Codon (ENC): Effective Number of Codons (ENC) is a well-established and widely used simple measure to quantify the degree of codon usage bias of a gene. ENC values range from 20-61 (Wright, 1990). It is generally believed that an ENC value of a gene <35 is considered to possess a strong codon bias. The larger the extent of codon preference in a gene, the smaller the ENC value is. The predicted values of ENC were calculated through using the following equation:

$$\text{ENC} = 2 + \text{S} + \frac{29}{\text{S}^2 + (1-\text{S})^2}$$

where, s represents the given $(G+C)_3\%$ value. If a gene lies on or just below the curve of the predicted values, its codon choice is constrained only by a mutation bias.

The calculation of the Relative Synonymous Codon Usage (RSCU): The Relative Synonymous Codon Usage (RSCU) can eliminate the influence of amino acid composition on codon usage and directly reflect the usage characteristics of codon bias. RSCU value refers to the ratio between the usage frequency of one codon in gene and expected frequency in all the synonymous codon.

Table 1: Morbilliviruses genomes used in this study

Virus (species)	Abbreviated name	Accession	Strain	Length[a]	Total genes[b]	Host
Measles virus	MV	NC_001498.1	Ichinose-B95a	14169	6	Human
Rinderpest virus	RPV	NC_006296.2	Kabete O	14133	6	Cattle
Peste-des-Petits-Ruminants virus	PPRV	NC_006383.2	Turkey 2000	14139	6	Ovis
Canine Distemper virus	CDV	NC_001921.1	Onderstpoort	14463	6	Canine
Dolphin Morbillivirus	CeMV	NC_005283.1	Undecided	14127	6	Dolphin

[a]The length values were excluding non-coding sequence and the units was bp; [b]The genes were excluding the non-structural protein genes

If the synonymous codon usage of one amino acid has no preferences or this codon is chosen equally and randomly in other words, the codon usage frequency is close to the expected frequency, then the RSCU value of this codon is equal to 1 if a codon RSCU value is >1, it indicates that the codon use frequency is higher than the expected frequency, otherwise, it is less than expected value (Sharp and Li, 1986).

The RSCU values in each gene of the Morbilliviruses were calculated according to the following calculating formula. In the formula, g_{ij} is the observed number of the ith codon for jth amino acid which has ni type of synonymous codons as:

$$\text{RSCU} = \frac{g_{ij}}{\sum_{j}^{n_i} g_{ij}} n_i$$

Statistical analysis: Correspondence Analysis (COA) is a mainly used statistical method to investigate the major trends in variation of codon usage in genes. It is used to plot all of the genes according to their RSCU values, taking them in a 59-dimensional hyperspace and each gene is described with 59 variables. The results would find out the major factors that affect the codon usage bias in each gene. In this study, correspondence analysis on Relative Synonymous Codon Usage (RSCU) was performed.

Correlation analysis was used to study the relationship between nucleotide composition and synonymous codon usage pattern among genes in Morbilliviruses. It was performed based on the Pearson's rank correlation analysis way.

The statistical analyzes were carried out with Statistical Software SPSS Version 19.0.

RESULTS AND DISCUSSION

The characteristics of synonymous codon usage in Morbilliviruses: In this study, detailed analyses were performed on the codon usage, nucleotide composition and other parameters that affect the codon usage in Morbilliviruses. The effective number of codons or ENC value among all the structural genes of Morbilliviruses were very similar and varied from 49.9-60.28 with a mean value of 54.85 and SD of 2.17 (Table 2). It suggested that the extent of codon preference in Morbilliviruses genomes is less biased (mean ENC>35) and keeps at a stable level.

RSCU values of 59 synonymous codons of each gene in Morbilliviruses are listed in a table.It seemed that major variation in RSCU is obvious among different genes and viruses (Fig. 1a and b). As a whole of the Morbilliviruses, some synonymous codons are chosen preferentially. They are ATC for Ile, GTC for Val, TCA for Ser, CCA for Pro, GCA for Ala, GAT for Asp, GAG for Glu, AGA and AGG for Arg. Some codons are rarely used or have low RSCU values. They are GTA for Val, TCG for Ser, CCG for Pro, CGT, CGC, CGA, CGG for Arg, GGC for Gly. And also, some codons for the same amino acid are chosen mainly equally. In other words, the codons for these amino acids are less biased. They are Phe, Leu, Tyr, His, Gln, Asn, Lys and Cys.

Overall, it was a consensus among the 6 curves of the codon usage in each gene group (Fig. 1a). However, there were several synonymous codons with discrepancy. The first quarter of the curve for *L* gene was smooth which indicated that each codon is used equally. There was a seemingly random variation in RSCU for several amino acids to individual virus (Fig. 1b). The coincident peaks of the curves implied that some codons are used with the highest frequency in all the ORFs of the Morbilliviruses.

Synonymous codon usage patterns of genomes in Morbilliviruses: Principal component analysis was used to analyze all the selected coding genes of viral genome in Morbilliviruses, taking them in a multidimensional hyberspace according their RSCU values. The 1st axis and 2nd represented the values of the first and the second axis of each gene in PCA. One major trend in the first axis accounted for 52.11% of the total variation and the second dimensional variable reflected 6.75% of the variation in PCA. It could be seen from the figure that each gene in Morbilliviruses was prominently located in different positions indicating that the codon usage patterns in Morbilliviruses are different (Fig. 2).

The relationship between nucleotide composition and codon usage in Morbilliviruses: The mean GC composition among Morbilliviruses ranged from 43.8- 47.79% while GC3 contents ranged from 40.73-48.99%. Thus, it was obvious that the average nucleotide composition of both the coding regions and synonymous positions of the codons keep at a stable level with a little huge range for the third synonymous position (Table 2). It is interesting that the mean GC composition of MV, RPV and PPRV were 47.7867, 47.5633, 47.2367%, respectively and CDV, CeMV were 43.88, 43.8%. So, the first three are in a level and the last two in another and the same with the GC3s contents. The results were consistent with the phylogenetic tree (Fig. 3) that had been built with the whole genomes of Morbilliviruses.

Table 2: Identified ENC and composition of the genes in the morbilliviruses genomes

Genome	Length	A (%)	C (%)	T (%)	G (%)	A3 (%)	C3 (%)	T3 (%)	G3 (%)	G+C (%)	(G+C)12 (%)	G3s+C3s (%)	ENC
N													
DMV	1572	30.22	21.37	23.35	25.06	29.96	22.90	25.57	21.56	46.44	47.425	42.35	54.99
MV	1578	29.15	21.48	22.69	26.68	28.71	23.38	23.19	24.71	48.16	48.195	45.94	53.73
PPRV	1578	27.63	23.19	21.48	27.69	25.29	27.76	19.39	27.57	50.89	48.670	53.66	53.85
RPV	1578	29.53	22.69	22.56	25.22	26.05	27.00	23.95	23.00	47.91	46.860	48.12	55.66
CDV	1572	29.90	20.93	25.45	23.73	27.10	23.28	28.44	21.18	44.66	44.750	42.00	55.98
P													
DMV	1521	32.35	21.89	21.24	24.52	30.37	21.30	27.81	20.51	46.42	48.715	40.73	54.60
MV	1524	28.94	25.46	19.42	26.18	26.18	29.13	22.05	22.64	51.64	51.575	50.51	55.07
PPRV	1530	30.92	23.99	20.65	24.44	26.67	28.63	23.92	20.78	48.43	47.940	49.01	55.03
RPV	1524	28.48	24.80	20.34	26.38	23.23	27.17	24.41	25.20	51.18	50.590	51.01	58.71
CDV	1524	32.48	21.85	20.47	25.20	29.92	21.06	28.54	20.47	47.05	49.805	40.32	53.56
M													
DMV	1008	29.56	20.14	26.49	23.81	24.40	24.40	28.87	22.32	43.95	42.560	44.03	49.90
MV	1008	28.47	23.31	23.21	25.00	22.92	29.76	22.32	25.00	48.31	45.090	52.80	51.09
PPRV	1008	29.96	22.22	23.81	24.01	28.87	27.98	22.32	20.83	46.23	44.940	46.58	53.90
RPV	1008	28.67	21.83	23.61	25.89	25.30	25.89	21.73	27.08	47.72	45.090	51.54	60.28
CDV	1008	30.46	19.94	26.79	22.82	30.06	24.11	27.08	18.75	42.76	42.710	40.50	58.57
F													
DMV	1659	31.10	20.25	26.52	22.12	33.27	20.80	27.49	18.44	42.37	43.940	37.73	54.69
MV	1653	28.01	22.38	24.14	25.47	25.77	25.59	22.14	26.50	47.85	45.735	50.93	56.87
PPRV	1641	29.92	22.43	23.58	24.07	29.07	24.50	22.49	23.95	46.50	45.520	47.49	52.89
RPV	1641	28.82	22.97	24.56	23.64	24.86	25.59	26.69	22.85	46.62	45.705	47.39	57.66
CDV	1989	29.86	23.53	25.34	21.27	27.90	24.74	27.60	19.76	44.80	44.950	43.19	55.35
H													
DMV	1815	30.96	19.83	26.39	22.81	28.26	22.98	27.77	20.99	42.64	41.985	41.95	52.32
MV	1854	27.83	23.03	25.30	23.84	23.95	26.86	25.08	24.11	46.87	44.825	48.82	55.20
PPRV	1830	27.10	22.13	26.61	24.15	23.44	25.25	28.20	23.11	46.28	45.245	46.78	55.81
RPV	1830	28.47	23.61	24.70	23.22	25.25	27.54	25.08	22.13	46.83	45.410	48.05	54.10
CDV	1815	29.81	21.21	27.49	21.49	29.92	22.15	28.43	19.50	42.70	43.220	39.73	53.66
L													
DMV	6552	31.07	19.20	27.95	21.78	28.34	19.41	30.86	21.38	40.98	41.070	38.59	53.44
MV	6552	29.76	21.37	26.34	22.53	26.14	24.18	28.02	21.66	43.89	42.925	43.87	54.43
PPRV	6552	29.01	22.08	25.90	23.00	24.18	26.60	26.24	22.99	45.09	42.835	47.77	56.36
RPV	6552	29.44	22.04	25.44	23.08	24.95	25.55	25.50	23.99	45.12	42.900	47.80	54.68
CDV	6555	30.65	20.26	28.04	21.05	28.70	21.24	30.62	19.45	41.31	41.625	38.63	53.25

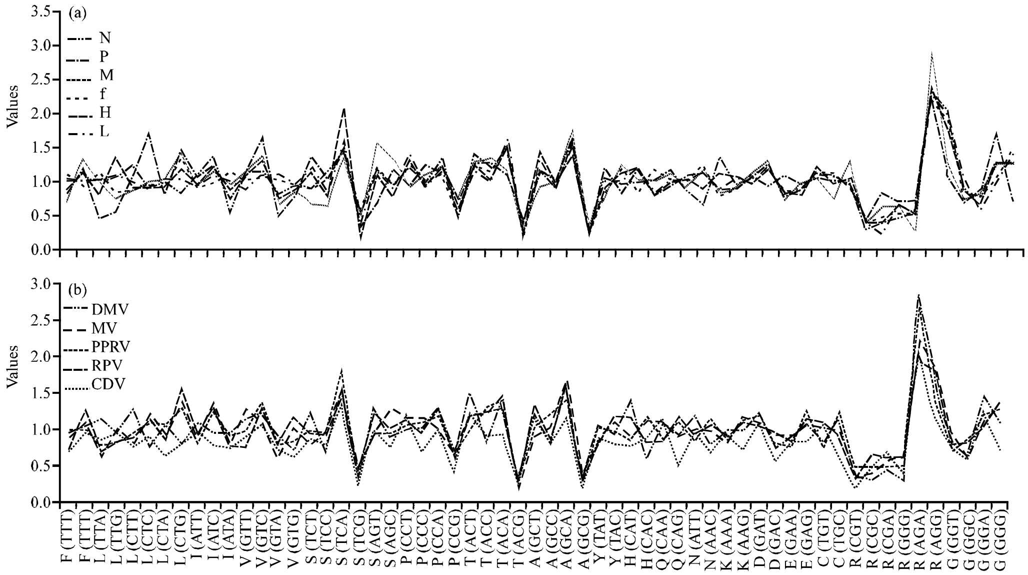

Fig. 1: a) Comparison the codon preferences among different ORFs of Morbilliviruses; b) Comparison the codon preferences among different viruses of Morbilliviruses

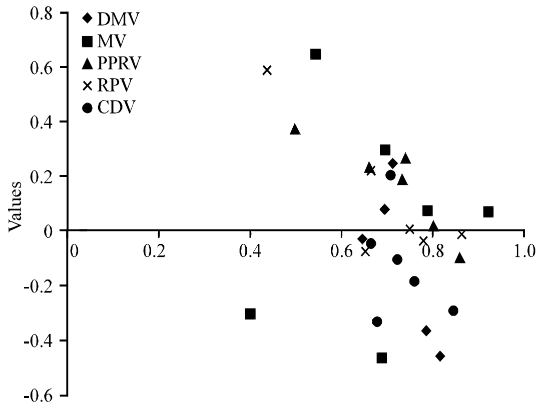

Fig. 2: A plot of the value of the first axis and the second axis of each gene of different viruses of Morbilliviruses in PCA

A strong and significant linear correlation (r = 0.822, $p<0.01$) was observed between the nucleotide composition (GC%) and nucleotide composition at only the third position of the codon (GC3s%) when considering all the 5 Morbilliviruses together. It proved that the 5 viruses of the same genus share the similar nucleotide composition patterns in their cording regions and they are phylogenetically close to each other. Furthermore, both (C+G) and (C+G)3s% had a highly significant correlation with each of A, T, C, G, A_3, C_3, G_3 and T_3%, respectively (Table 3).

And taking the result of correlation of GC and GC3s% into account indicated that GC content at only the third position of the codon affects the interaction between mutation pressure and natural selection to a certain extent and also showed that nucleotide composition restriction affects the codon usage pattern of Morbilliviruses.

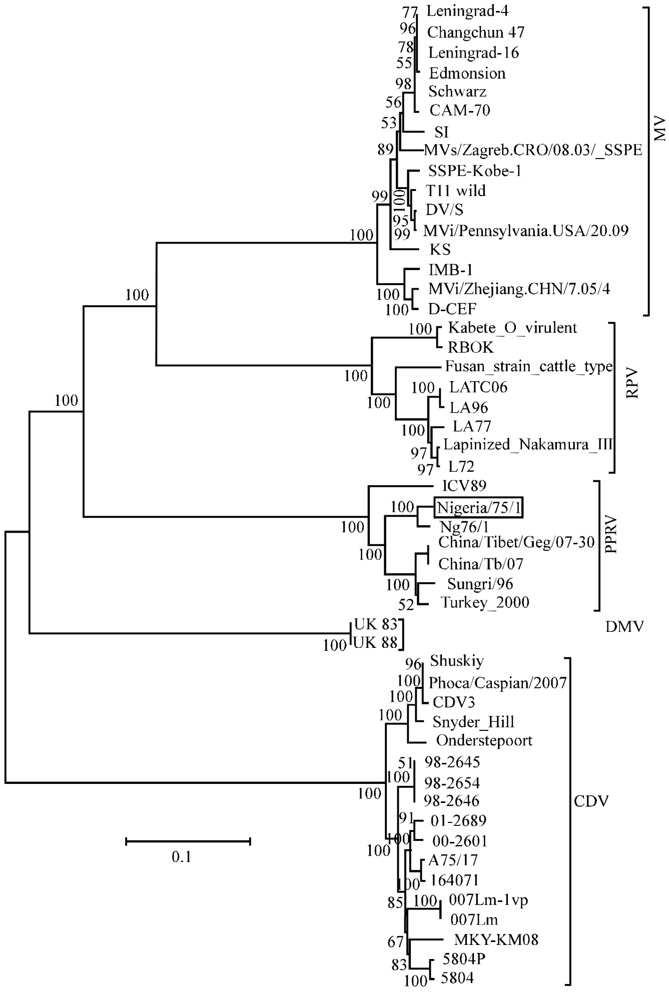

Fig. 3: The phylogenetic tree of the different viruses in Morbilliviruses. The tree was built by neighbor-joining method in MEGA5 with all the whole genome sequences of Morbilliviruses that were available in NCBI

Table 3: Correlation analysis between the (C+G)%, (C+G)3s% and A%, T%, C%, G%, A3%, C3%, G3%, T3% in the cording region of the Morbilliviruses genomes

Correlation analysis	A (%)	G (%)	T (%)	C (%)	A_3 (%)	G_3 (%)	T_3 (%)	C_3 (%)
(C+G) %								
r	0.478	0.885	0.882	0.852	0.468	0.672	0.798	0.714
p	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
$(C+G)_{3s}$%								
r	0.752	0.686	0.533	0.747	0.762	0.85	0.884	0.888
p	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Furthermore, the linear regression analysis between synonymous codon usage and nucleotide composition was applied. It showed that the Axis1a of each gene calculated by PCA is closely correlated with the $(G+C)_{12}$% (r = 0.738, p<0.01) and $(C+G)_{3s}$% (r = 0.485, p<0.01). Finally, the ENC-plot (ENC plotted against GC3s%) of each gene and individual virus were constructed on graphs and were used as a part of general strategy to investigate patterns of synonymous codon usage. All the plots of genes were well located or on the lower part of the expected curve and all the plots of 5 viruses lied below the expected curve (Fig. 4a and b). The nucleotide composition at all three positions in a codon triplet differs, since these positions are subject to different selective constraints.

The base content of the first and the second codon position is considered to be determined by selective constraints while the content in the third synonymous position is supposed to be largely determined by mutational pressures. The diagrams suggested that codon usage of these genes has pertinence with mutation bias rather than natural selection. The same literature about RNA viruses had been reported previously (Drake and Holland, 1999). It is generally believed that the mutation rate of RNA virus genome is much higher than the DNA viruses (Jenkins and Holmes, 2003).

Preferences for particular dinucleotides were observed in Morbilliviruses genera (Fig. 5). It can be seen that most dinucleotides were randomly distributed. The dinucleotides CG were appreciably underrepresented in the Morbilliviruses. The RSCU values of eight codons containing TCG, CCG, ACG, GCG, CGT, CGC, CGA and CGG were lower. All of these eight codons were not preferential codons. It may be the reason of the low frequency of dinucleotide CG.

It was observed that the 5 curves can be divided into 2 groups. For CDV and DMV, the frequency of dinucleotides AT/TA, TT and AA were higher than the others and the CC, GG and GC were lower. On the basis of the observations, it was clear that the Morbilliviruses can be divided into two groups based on patterns of dinucleotide frequency usage.

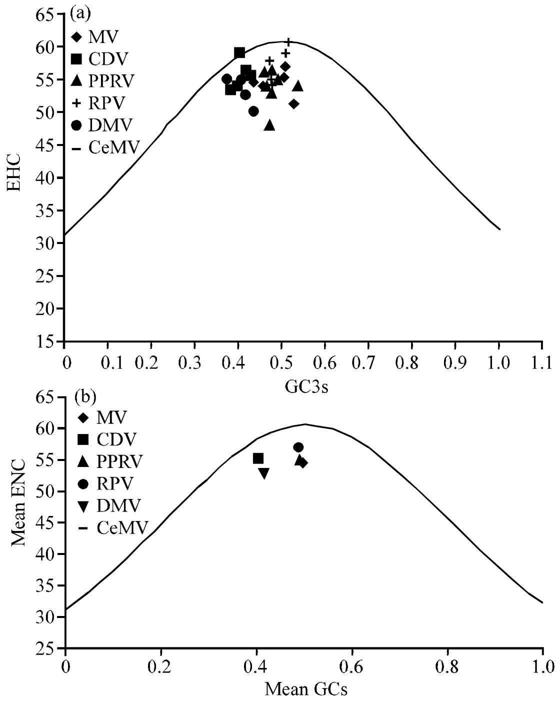

Fig. 4: a) ENC values and GC3S plot for all *Morbilliviruses* genes; b) ENC values and GC3S plot for individual Morbilliviruses

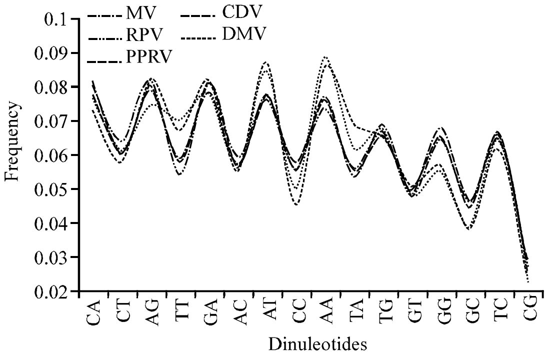

Fig. 5: Dinucleotide frequency of the 5 viruses of Morbilliviruses

In this study, the ENC value of Morbilliviruses, researchers calculated was 54.85. So, the synonymous codon usage bias in Morbilliviruses is unconspicuous. It is agree with the other RNA viruses that earlier reported such as BVDV, H5N1 influenza virus and SARS-covs (Gu *et al*., 2004a; Wang *et al*., 2011). For the Pathogenic microorganism, it is advantageous to the multiplication for themselves in host cells.

It was strongly considered that codon usage pattern formation is mainly caused by the mutation pressure rather than the natural selection. According to the results of the sequences in Morbilliviruses, researchers showed that the influence of mutation pressure is greater than the natural selection.

In earlier reports, the codon usage patterns and gene length for some viruses had significant correlations and there were some insignificant correlations (RoyChoudhury and Mukherjee, 2010). In the study, researchers found that the length has no significant effect on the codon usage of the coding genes in the Morbilliviruses (p = -0.09).

Relative abundance dinucleotides analysis presented that the frequency of CpG is significant low at all positions for coding genes in 5 Morbilliviruses. Figure 1a and b showed that the RSCU values of eight codons containing TCG, CCG, ACG, GCG, CGT, CGC, CGA and CGG are lower. In the earlier report, the same feature was existed in NDV genome (Wang *et al*., 2009). So, it may be a common characteristic among all the members in Paramyxoviridae.

From the plot in PCA of different virus of Morbilliviruses, researchers concluded that the codon usage patterns of Morbilliviruses are different. Dinucleotide frequency of the different viruses of Morbilliviruses revealed that the 5 viruses could be divided into 2 groups, MV, RPV and PPRV in a group, CDV, CeMV in another one. ENC values and GC_{3S} plot for individual Morbilliviruses from Fig. 4b also supported this viewpoint.

CONCLUSION

The analysis revealed that the codon usage bias in Morbilliviruses is low and mutational pressure is the main factor that affects the variation in this genus. The codon usage patterns of the five members in Morbilliviruses are different and they can be divided into two groups. Researchers also, found that CpG under-represented is a characteristic in Morbilliviruses. Researchers speculated that it may be a common feature among the Paramyxoviridae. The results not only gave an insight into the variation of codon usage pattern among the genomes in Morbilliviruses but also played an important role for understanding the evolution process of the viruses in Morbilliviruses.

ACKNOWLEDGEMENTS

This study was financially supported by the Special Fund for Agro-Scientific Research in the Public Interest (No. 201103008) and the earmarked fund for China Agriculture Research System (No. CARS-40-10).

REFERENCES

Archetti, M., 2004. Codon usage bias and mutation constraints reduce the level of error minimization of the genetic code. J. Mol. Evol., 59: 258-266.

Bellini, W.J., G. Englund, S. Rozenblatt, H. Arnheiter and C.D. Richardson, 1985. Measles virus P gene codes for two proteins. J. Virol., 53: 908-919.

Bulmer, M., 1988. Codon usage and intragenic position. J. Theoretical Biol., 133: 67-71.

Carbone, A., A. Zinovyev and F. Kepes, 2003. Codon adaptation index as a measure of dominating codon bias. Bioinformatics, 19: 2005-2015.

Cattaneo, R., K. Kaelin, K. Baczko and M.A. Billeter, 1989. Measles virus editing provides an additional cysteine-rich protein. Cell, 56: 759-764.

Diallo, A., 1990. Morbilliviruses group: Genome organisation and proteins. Vet. Microbiol., 23: 155-163.

Drake, J.W. and J.J. Holland, 1999. Mutation rates among RNA viruses. Proc. Nat. Acad. Sci., 96: 13910-13913.

Epstein, R.J., K. Lin and T.W. Tan, 2000. A functional significance for codon third bases. Gene, 245: 291-298.

Fu, M., 2010. Codon usage bias in herpesvirus. Arch. Virol., 155: 391-396.

Gu, W., T. Zhou, J. Ma, X. Sun and Z. Lu, 2004a. Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. Virus Res., 101: 155-161.

Gu, W., T. Zhou, J. Ma, X. Sun and Z. Lu, 2004b. The relationship between synonymous codon usage and protein structure in *Escherichia coli* and Homo sapiens. Biosystems, 73: 89-97.

Jenkins, G.M. and E.C. Holmes, 2003. The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res., 92: 1-7.

Kanaya, S., M. Kinouchi, T. Abe, Y. Kudo and Y. Yamada *et al*., 2001. Analysis of codon usage diversity of bacterial genes with a Self-Organizing Map (SOM): Characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome. Gene, 276: 89-99.

Karlin, S. and J. Mrazek, 1996. What drives codon choices in human genes? J. Mol. Biol., 262: 459-472.

Karlin, S., L. Brocchieri, J. Trent, B.E. Blaisdell and J. Mrazek, 2002. Heterogeneity of genome and proteome content in bacteria, archaea and eukaryotes. Theoretical Populat. Biol., 61: 367-390.

Liu, X., C. Wu and A.Y. Chen, 2010. Codon usage bias and recombination events for neuraminidase and hemagglutinin genes in Chinese isolates of influenza A virus subtype H9N2. Arch. Virol., 155: 685-693.

Liu, X.S., Y.L. Wang, Y.G. Zhang, Y.Z. Fang and L. Pan *et al*., 2011. Analysis of codon usage in peste des petits ruminant's virus. Afr. J. Microbiol. Res., 5: 4592-4600.

RoyChoudhury, S. and D. Mukherjee, 2010. A detailed comparative analysis on the overall codon usage pattern in herpesviruses. Virus Res., 148: 31-43.

Roychoudhury, S., A. Pan and D. Mukherjee, 2010. Genus specific evolution of codon usage and nucleotide compositional traits of poxviruses. Virus Genes, 42: 189-199.

Shackelton, L.A., C.R. Parrish and E.C. Holmes, 2006. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J. Mol. Evolut., 62: 551-563.

Sharp, P.M. and W.H. Li, 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for rarecodons. Nucleic Acids Res., 14: 7737-7749.

Sharp, P.M., E. Cowe, D.G. Higgins, D.C. Shields, K.H. Wolfe and F. Wright, 1988. Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*: A review of the considerable within-species diversity. Nucleic Acids Res., 16: 8207-8211.

Supek, F. and K. Vlahovicek, 2005. Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity. BMC Bioinformatics, Vol. 6.

Tao, P., L. Dai, M. Luo, F. Tang, P. Tien and Z. Pan, 2009. Analysis of synonymous codon usage in classical swine fever virus. Virus Genes, 38: 104-112.

Wang, M., J. Zhang, J.H. Zhou, H.T. Chen and L.N. Ma *et al*., 2011. Analysis of codon usage in bovine viral diarrhea virus. Arch. Virol., 156: 153-160.

Wang, M., Y. Liu, J. Zhou, H. Chen and L. Ma *et al*., 2009. Analysis of codon usage in Newcastle disease virus. Virus Genes, 42: 245-253.

Wright, F., 1990. The effective number of codons used in a gene. Gene, 87: 23-29.

Zhong, J., Y. Li, S. Zhao, S. Liu and Z. Zhang, 2007. Mutation pressure shapes codon usage in the GC-Rich genome of foot-and-mouth disease virus. Virus Genes, 35: 767-776.