Comparative characterization analysis of synonymous codon usage bias in classical swine fever virus

CrossMark

Xin Xu [a, b, c], Dongliang Fei [a], Huansheng Han [a], Honggui Liu [a], Jiayong Zhang [a], Yulong Zhou [d, e], Chuang Xu [d, e], Hongbin Wang [a, *], Hongwei Cao [c, d, e, **], Hua Zhang [c, d, e, ***]

[a] College of Animal Medicine, Northeast Agricultural University, Harbin 150030, China
[b] Heilongjiang Institute of Veterinary Science, Qiqihar 161005, China
[c] College of Life Science and Technology, HeiLongJiang BaYi Agricultural University, Daqing 163319, China
[d] College of Animal Science and Veterinary Medicine, HeiLongJiang BaYi Agricultural University, Daqing 163319, China
[e] Biotechnology Center, HeiLongJiang BaYi Agricultural University, Daqing 163319, China

ARTICLE INFO

ABSTRACT

Classical swine fever virus (CSFV) is responsible for the highly contagious viral disease of swine, and causes great economic loss in the swine-raising industry. Considering the significance of CSFV, a systemic analysis was performed to study its codon usage patterns. In this study, using the complete genome sequences of 76 CSFV representing three genotypes, we firstly analyzed the relative nucleotide composition, effective number of codon (ENC) and synonymous codon usage in CSFV genomes. The results showed that CSFV is GC-moderate genome and the third-ended codons are not preferentially used. Every ENC values in CSFV genomes are >50, indicating that the codon usage bias is comparatively slight. Subsequently, we performed the correspondence analysis (COA) to investigate synonymous codon usage variation among all of the CSFV genomes. We found that codon usage bias in these CSFV genomes is greatly influenced by G + C mutation, which suggests that mutational pressure may be the main factor determining the codon usage biases. Moreover, most of the codon usage bias among different CSFV ORFs is directly related to the nucleotide composition. Other factors, such as hydrophobicity and aromaticity, also influence the codon usage variation among CSFV genomes. Our study represents the most comprehensive analysis of codon usage patterns in CSFV genome and provides a basic understanding of the mechanisms for its codon usage bias.

© 2017 Elsevier Ltd. All rights reserved.

1. Short communication

Classical swine fever (CSF) is firstly recognized in Tennessee, USA in 1810, and is described in France in 1822 [1]. As one of the office international des epizooties (OIE) notifiable diseases, CSF caused significant economic losses in the swine-raising industry worldwide [2]. CSF is an extremely contagious swine disease with high morbidity and mortality, featuring symptoms of hemorrhagic fever and immuno-suppression, which is caused by classical swine fever virus (CSFV) [3]. CSFV is a member of the genus *Pestivirus*

within the family *Flaviviridae*, which is an enveloped virus harboring a single strand positive-sense RNA genome with approximately 12, 300 nucleotides in length [4]. The genome of CSFV, comprising a single long open reading frame (ORF) that encodes a polyprotein composed of 3898 amino acids (aa), flanked by two non-coding regions at the 3′ untranslated region (3′-UTR) and 5′ untranslated region (5′-UTR) [5]. The polyprotein is subsequently processed into twelve mature proteins by viral and cellular proteases, including four structural proteins (C, E^{rns}, E1 and E2) and eight nonstructural proteins (N^{pro}, P7, NS2, NS3, NS4A, NS4B, NS5A

and NS5B) [6].

The CSFVs are classified into highly virulent, moderately virulent, lowly virulent, and avirulent strains. Phylogenetic analysis is extensively used for tracing CSFV and analyzing its epidemiological situation [7]. Based on sequence data-sets of the envelope glycoprotein gene (E2), polymerase gene (NS5B) and untranslated region (5′-UTR), phylogenetic analysis divides CSFVs into three genotypes, 1, 2, and 3, with each being further divided into three or four subgenotypes [8]. Recently, several studies reported that vaccination might affect CSFV diversity and immune escape through recombination and point mutation. At the same time, vaccination may influence the population dynamics, evolutionary rate and adaptive evolution of CSFV [1]. It is well known that synonymous codons are not used randomly. Codon usage is also found to be related to codon-anticodon interaction, dinucleotide bias, tRNA abundance, gene length, gene function, protein secondary structure, replicational and translational selection, and tissue or organ specificity [9,10]. Mutational pressure and translational selection are thought to be the main factors that account for codon usage variation among genes in some RNA virus [11,12]. Therefore, it is essential to the understanding of viral evolution, particularly the interplay between viruses and the host immune response through studying the extent and causes of biases in codon usage [13]. Previous studies of CSFV have mainly been limited to phylogenic analysis, and few synonymous codon usage analyses have been applied. In order to better understand the characteristics of the CSFV genome and to reveal more information about the viral genome, the systemic analysis was performed to study its codon usage patterns. In addition, spearman's rank correlation analysis was used to determine the role of different factors in shaping the codon usage biases in the various CSFV genomes. All statistical analyses were carried out using the statistical analysis software SPSS (Version 17.0).

In this present study, we firstly sought to address the issues concerning codon usage in CSFV genome. A total of 76 publicly available complete CSFV genomes representing three genotypes isolated from all over the world were obtained from Genebank (http://ncbi.nlm.nih.gov). The sequences with >99% sequence identities were excluded. The GenBank accession numbers and other detail information of each CSFV genome are listed in Table 1. Relative synonymous codon usage (RSCU) values are largely independent of amino acid composition and are particularly useful in comparing codon usage between genes, or sets of genes that differ in their size and amino acid composition [14]. For the sake of examining synonymous codon usage without the confounding influence of amino acid composition of different CSFV genome, RSCU values of each codon in each ORF were used to measure the synonymous codon usage. The preferentially used codons are A-ended (4 ones), U-ended (1 ones), C-ended (8 ones) and G-ended (6 ones) codons (Table 2). The average GC content of all CSFV genome is 46.38% (From 45.42% to 47.23%, with a standard deviation (S.D.) of 0.44%), and the average third position content in synonymous

Table 1
List of CSFV strains used for analysis of synonymous codon usage in this study.

Genebank accession	GC3s	ENC	Mononucleotide frequencies				Genebank accession	GC3s	ENC	Mononucleotide frequencies			
			C	T	A	G				C	T	A	G
AY259122	0.5023	52.0916	0.1987	0.2212	0.3060	0.2632	GU233732	0.5145	52.4512	0.2039	0.2165	0.3068	0.2624
KT119352	0.5182	51.1479	0.2060	0.2135	0.3079	0.2627	GU233731	0.5149	52.3199	0.2031	0.2172	0.3069	0.2623
KP233071	0.5153	51.6138	0.2063	0.2140	0.3086	0.2608	AY367767	0.5156	51.3559	0.2045	0.2154	0.3079	0.2620
KF977610	0.4955	51.9976	0.1981	0.2215	0.3098	0.2600	AY646427	0.5114	52.3832	0.2012	0.2190	0.3079	0.2622
KF977609	0.4955	51.9896	0.1981	0.2215	0.3099	0.2599	DQ127910	0.4947	51.7132	0.1972	0.2214	0.3105	0.2601
KF977608	0.4956	51.9890	0.1983	0.2213	0.3100	0.2598	HQ148063	0.5143	51.4004	0.2044	0.2150	0.3086	0.2615
KF977607	0.4959	51.9850	0.1984	0.2213	0.3097	0.2599	HQ148062	0.5151	52.3274	0.2057	0.2150	0.3074	0.2612
HQ380231	0.4933	51.7254	0.1972	0.2219	0.3106	0.2594	HQ148061	0.5143	52.2577	0.2053	0.2151	0.3086	0.2611
AY775178	0.4938	51.7114	0.1971	0.2219	0.3102	0.2599	HM175885	0.5072	52.0137	0.1989	0.2211	0.3046	0.2646
KU504339	0.5286	51.2851	0.2072	0.2118	0.3067	0.2644	HM237795	0.4966	51.9944	0.1984	0.2213	0.3095	0.2601
KU556758	0.5235	51.3696	0.2067	0.2120	0.3070	0.2633	X87939	0.4973	52.0328	0.1979	0.2216	0.3092	0.2603
KT716271	0.5151	52.0706	0.2019	0.2165	0.3062	0.2660	AY578688	0.4995	52.2760	0.2009	0.2183	0.3107	0.2594
KF669877	0.5200	51.6150	0.2021	0.2158	0.3068	0.2659	AY578687	0.4950	51.6731	0.2008	0.2189	0.3115	0.2578
KP233070	0.5087	51.2914	0.2025	0.2160	0.3082	0.2633	AY663656	0.5069	52.0515	0.1985	0.2216	0.3043	0.2646
KM362426	0.5326	51.3687	0.2094	0.2108	0.3071	0.2629	GQ902941	0.5210	51.7554	0.2060	0.2143	0.3071	0.2625
NC_002657	0.4975	51.8972	0.1988	0.2207	0.3102	0.2593	GQ122383	0.5206	50.9538	0.2079	0.2120	0.3081	0.2622
KM262189	0.4883	51.5693	0.1958	0.2240	0.3110	0.2584	AY554397	0.5239	51.4640	0.2061	0.2131	0.3074	0.2631
KJ619377	0.5218	52.1422	0.2056	0.2141	0.3068	0.2629	AY568569	0.5209	51.2686	0.2065	0.2133	0.3078	0.2619
KC149991	0.5189	51.1642	0.2059	0.2141	0.3077	0.2620	J04358	0.5180	52.1828	0.2060	0.2139	0.3078	0.2619
KC149990	0.5187	51.3783	0.2047	0.2147	0.3074	0.2631	FJ265020	0.5109	52.3005	0.2043	0.2165	0.3075	0.2615
JX262391	0.5055	51.2231	0.2027	0.2159	0.3105	0.2606	EU497410	0.4952	51.6385	0.1973	0.2215	0.3107	0.2596
JX218094	0.5081	51.2548	0.2032	0.2153	0.3102	0.2610	LT158502	0.5107	52.0471	0.2049	0.2154	0.3082	0.2612
GU592790	0.5175	51.5211	0.2062	0.2124	0.3078	0.2630	LT158410	0.5107	52.0471	0.2049	0.2154	0.3082	0.2612
AY382481	0.5083	52.1012	0.1992	0.2210	0.3044	0.2646	LT158409	0.5073	52.0830	0.2046	0.2156	0.3090	0.2603
AF326963	0.4975	51.8972	0.1988	0.2207	0.3102	0.2593	LT158408	0.5064	52.0768	0.2046	0.2157	0.3090	0.2602
AY805221	0.5063	52.0463	0.1988	0.2211	0.3048	0.2645	LT158407	0.5089	52.1037	0.2047	0.2156	0.3088	0.2605
GQ923951	0.5161	51.9421	0.2030	0.2158	0.3081	0.2620	LT158406	0.5069	52.0596	0.2044	0.2161	0.3088	0.2603
EU789580	0.4955	52.1820	0.1977	0.2214	0.3092	0.2607	LT158405	0.5079	52.0196	0.2046	0.2158	0.3089	0.2604
FJ529205	0.5173	51.2519	0.2057	0.2134	0.3062	0.2642	LT158404	0.5086	52.0139	0.2048	0.2156	0.3089	0.2603
EU857642	0.5052	52.2717	0.1985	0.2216	0.3037	0.2656	LT158403	0.5100	52.1054	0.2047	0.2158	0.3081	0.2611
EU490425	0.4947	51.9719	0.1977	0.2219	0.3094	0.2601	LT158402	0.5112	52.1138	0.2049	0.2158	0.3077	0.2613
KP343640	0.5131	51.1766	0.2046	0.2140	0.3091	0.2623	LT158401	0.5101	52.0425	0.2049	0.2156	0.3086	0.2607
KC503764	0.4880	51.6418	0.1965	0.2238	0.3111	0.2577	KJ873238	0.4975	51.9234	0.1993	0.2194	0.3104	0.2601
KC851953	0.5061	52.5419	0.2032	0.2164	0.3098	0.2588	KM522833	0.4983	52.1057	0.2003	0.2186	0.3111	0.2593
EU915211	0.4953	52.1666	0.1977	0.2215	0.3093	0.2605	JQ268754	0.5178	51.2389	0.2051	0.2143	0.3078	0.2624
GU324242	0.5093	52.0929	0.2038	0.2160	0.3083	0.2615	AF531433	0.5069	52.0395	0.1988	0.2212	0.3046	0.2646
GU233734	0.5151	52.5616	0.2032	0.2170	0.3065	0.2627	AF407339	0.5058	51.3742	0.2041	0.2161	0.3089	0.2603
GU233733	0.5160	52.5102	0.2035	0.2168	0.3062	0.2630	AF333000	0.4941	51.6170	0.1973	0.2217	0.3103	0.2599