Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Senecavirus A evolution

CrossMark

Ye Chen [a, 1], Quanming Xu [a, b, 1], Chen Tan [a], Xinxin Li [a], Xiaojuan Chi [a], Binxiang Cai [a], Ziding Yu [a], Yanmei Ma [a, **], Ji-Long Chen [a, c, *]

[a] *Key Laboratory of Fujian-Taiwan Animal Pathogen Biology, College of Animal Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China*
[b] *College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou, 350002, China*
[c] *CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China*

ARTICLE INFO

ABSTRACT

Senecavirus A (SVA) infection was recently confirmed in pigs in Brazil, United States of America and Canada. To better understand the molecular characteristics of isolated SVA genomes, we first reported genome-wide comprehensive analyses of codon usage and various factors that have contribute to the molecular evolution in SVA. The effective number of codons (ENC) ranged from 54.51 to 55.54 with an average of 54.87 ± 0.285, which reveals a relatively stable nucleotide composition. We found that codon usage bias of the SVA was low. Mutational pressure acted as an increasingly dominant factor for the evolution of the virus compared with the natural selection. Notably, codon usage bias was also affected by the geographic distribution and isolated time. The first systemic analysis on the codon usage bias of the SVA provides important information for the understanding of the evolution of the SVA and has fundamental and theoretical benefits.

© 2017 Published by Elsevier Ltd.

1. Background

Senecavirus A (SVA), commonly known as Seneca Valley virus (SVV), is the only member of the genus Senecavirus within the family Picornaviridae. SVA is a single-stranded, positive-sense, non-enveloped RNA virus with an approximately 7.2 kb genome [1]. This virus was first discovered as a contaminant in 2001 (and named Seneca Valley virus 001 [SVV-001]) while cultivating viral vectors in the PER.C6 cell line [2,3]. While, SVA was isolated in previous cases collected from various pig farms in the USA from 1988 to 2001 [4]. One polyprotein of SVA is post-translationally processed by virus-encoded proteases into 4 structural (VP1-4) and 7 non-structural (2A-2C, 3A-3D) proteins [2,5], among which VP1 is considered to be the most immunogenic protein in the family Picornaviridae [6,7].

The main clinical symptoms of animals infected with SVA were vesicles on coronary bands or the snouts, sometimes exhibited acute lameness, anorexia, lethargy, and transient fever. Infected breeding herds had an increase of neonatal morbidity and mortality ranging from 30% to 70%, mainly piglets less than 7-day-old [8,9]. The clinical symptoms of SVA resembles foot-and-mouth disease, swine vesicular disease, vesicular exanthema of swine, and vesicular stomatitis, which are four vesicular diseases [10]. SVA was also found in lesions in pigs suffering from porcine idiopathic vesicular disease in Canada and USA in 2008 and 2012, respectively [11]. In 2014 and 2015, SVA infection was associated with outbreaks of vesicular disease in sows as well as neonatal pig mortality in Brazil and USA [12]. In China, SVA (SVV CH-01-2015) was also first isolated in a pig farm in 2015 [13].

Except for methionine and tryptophan, other amino acids can be coded by more than one codon due to redundancy in the genetic code, also known as synonymous codon usage. However, the usage of various codons to code amino acids is not random and some are used more often, which is known as codon usage bias [14]. Codon usage bias has been reported for some RNA viruses, but the rate can vary depending on the identity of the virus. For instance, rubella

* Corresponding author. Key Laboratory of Fujian-Taiwan Animal Pathogen Biology, College of Animal Sciences, Fujian Agriculture and Forestry University, Fuzhou, 350002, China.
** Corresponding author.
E-mail addresses: 89mym@163.com (Y. Ma), chenjl@im.ac.cn (J.-L. Chen).
[1] These authors contributed equally to this work.

and rotavirus have strong codon usage bias, whereas porcine circovirus type 2 (PCV2) and porcine epidemic diarrhea virus (PEDV) have weak codon usage bias [15,16]. Natural selection, gene length, mutation pressure, abundance of tRNAs and RNA structure all affect codon usage bias. The relation of codon usage among viruses and their hosts is expected to affect viral survival, fitness, evasion from the host immune system and evolution [17].

Considering the recent increase in the worldwide prevalence of SVA and its potential risk for the pig industry, in this study, we first report genome-wide comprehensive analyses of codon usage and various factors that have contributed to the molecular evolution of SVA.

2. Materials and methods

2.1. Sequence data

In this study, 23 complete genome and complete coding sequences of SVA isolates were retrieved from the National Center for Biotechnology (NCBI) GenBank database (http://www.ncbi.nlm.nih.gov). To maintain the statistical significance of codon usage bias, artificial sequences were not included. Detailed information of the 23 strains, including the accession number, the location and date of isolation were listed in the supplemental materials (Table S1). The data set comprised of 14 complete genome sequences and 9 complete coding sequences.

3. Indices of codon usage and synonymous codon usage bias

3.1. Nucleotide composition analysis

The nucleotide content (A%, U%, G%, C%) of each SVA strain was calculated using BioEdit (version 7.0.9.0) software. Each nucleotide at the third position of the synonymous codons (A3%, U3%, G3%, C3%) was analyzed using Codon W program online (http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::CodonW). The G + C at the first (GC1s), second (GC2s) and third codon positions (GC3s) of each isolates were calculated using the cusp program online (http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp).

3.1.1. Relative synonymous codon usage (RSCU) analysis

The relative synonymous codon usage (RSCU), proposed in 1986 (Sharp and Li, 1986) is widely used to evaluate the codon usage bias between genes or sets of genes that differ in their size and amino acid composition [15]. The RSCU values are not confounded by amino acid composition, with the values are the ratio of its observed number to its standard number on all codons for a particular amino acid are used randomly [18]. It was calculated using the following equation:

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

where g_{ij} is the observed number of codons for the amino acid, which has ni kinds of synonymous codons. A higher RSCU value indicates a stronger codon usage bias. It's considered that codon is used equally when RSCU value is 1.0; if RSCU is more than 1.0, the codon usage bias is positive; if it is less than 1.0, the codon usage bias is considered to be negative. In addition, codons with RSCU values $\geq$ 1.6 are over-represented, and codons with RSCU values $\leq$ 0.6 are under-represented [19]. EMBOSS: cusp was used for the RSCU analysis. (http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp).

3.1.2. Effective number of codons(ENC) analysis

To quantify the magnitude of the codon usage bias of each gene, the ENC value of each strain was calculated. The ENC is the best estimator of absolute synonymous codon usage bias [20] and was calculated using the following formula:

$$ENC = 2 + \frac{9}{F_2} + \frac{1}{F_3} + \frac{5}{F_4} + \frac{3}{F_6}$$

where $F_{(i=2,3,4,6)}$ is the mean of F_i for the i-fold degenerate amino acids. The F_i values were calculated using the formula below:

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

where n is the total number of frequencies of the codons for that amino acid and n_j is the total number of frequencies of the codon for that amino acid. In contrast to the RSCU, a lower ENC value indicates a higher codon usage bias. Normally, the ENC values range from 20 to 61 [21].If only one of the possible synonymous codons is used for the corresponding amino acid, the ENC is 20. While there is no codon usage bias, the ENC value is 61 [21]. Therefore, if the ENC is equal to or less than 35, the codon usage bias is considered extremely strong [20,21].

To determine the major factors that affect the codon usage bias, an ENC-plot was generated in which the ENC was plotted against the GC3s. When the codon usage is only constrained by the GC3s, the observed ENC is on or close to the null model (standard curve). If other factors such as natural selection also play role in the codon usage pattern, the observed values are far below the standard curve [22]. The expected ENC was calculated using the equation:

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

where s is the frequency of G + C at the third codon position of synonymous codons.

3.1.3. General average hydropathicity (Gravy) and aromaticity (Aroma) indices analysis

To improve the understanding of the influence of natural selection on shaping the codon usage bias, the Gravy and Aroma scores were determined in this study. These indices were obtained from CodonW **1.4.4** (http://codonW.sourceforge.net//), which can signify the frequencies of hydrophobic and aromatic amino acids. Therefore, the variation of the two indices indicates the amino acid usage. A higher Gravy or Aroma value suggests a more hydrophobic or aromatic amino acid product, respectively.

3.1.4. Principal component analysis (PCA)

Principal component analysis (PCA), a multivariate statistical approach in codon usage analysis, was widely used to analyze the major trends in codon usage patterns among different SVA strains [23]. In the PCA, a 59-dimensional vector corresponds to the RSCU of each strain, excluding the codons of AUG, UGG and termination codons, which transform RSCU values into uncorrelated variables. The PCA combined with the correlation analysis effectively demonstrated the factors influencing the codon usage bias.

3.1.5. Neutral evolution analysis

To investigate the varying role of mutational pressure and natural selection in shaping the codon usage bias of porcine SVA, a neutrality plot was drawn using GC12s as the ordinate and GC3s as the abscissa [24]. In the neutrality plot, each dot represents an

independent SVA isolates. In general, if the linear regression has a slope of 1, the effect is completely neutral, whereas a slope of 0 is indicative of complete selective constraints [24]. In addition, the GC12s or the GC3s values are plotted against the date of isolation to estimate the evolutionary pattern of the porcine SVA. The slope of a simple regression line explains the evolutionary speed of the natural selection pressure and the mutational pressure, respectively.

3.1.6. Statistical analysis

Using the statistical software SPSS (version 22.0) and GraphPad Prism6.0 with one-way analysis of variance (ANOVA) methods, a correlation analysis was performed. The figures related to this analysis were drawn by GraphPad Prism 6.0.

4. Results

4.1. Nucleotide composition of the porcine SVA

In the present study, 23 sequences of porcine SVA, including 14 complete genome sequences and 9 complete coding sequences, were analyzed. The composition properties of the SVA strains are shown in Table 1. The results reveal that the mean C% was 28.59%, and the mean G% was 23.07%, which were the highest and lowest, respectively. The mean GC% composition was 51.66%. Thus, it appears that there may be more usage bias of C nucleotides among SVA codons. To gain further insight into the influence of nucleotide content in shaping the SVAs codon usage pattern, the nucleotide composition at the third codon position (A3, U3, G3, C3, GC3) were calculated and listed in Table 2, which suggests that the mean C3% (41.06%) is the highest. The GC3 fluctuated from 55.41% to 57.15% with a mean of 56.6%, indicating that G/C-ended codons may be preferred over A/U-ended codons in SVAs sequences.

4.2. Synonymous codon usage in porcine SVA

The RSCU of the overall 59 sense codons were calculated to explore the extent of the preferred G/C-ended codons, and are listed in Table 3. The results showed that the most frequently used codon is GAC (Asp, 1799 times), and the least frequently used codon is AGG (Arg, 186 times). Among the hydrophobic amino acids, the most frequently used codon is GAC (Asp, 1799 times), and the least frequently used codon is AGG (Arg, 186 times). Among the hydrophilic amino acids, the most frequently used codon is GCC (Ala, 1627 times), and the least frequently used codon is AUA (Ile, 186 times). Among the 18 most abundantly used codons, 14 codons were G/C-ended, and 4 codons were A/U-ended, which evidently confirmed that there exists high codon usage bias in SVA sequences.

Furthermore, it is quite interesting to note that mostly over-represented and under-represented codons are C-ended and A/U-ended, respectively, which also reveals that the preferred codons are influenced by compositional constraints.

Additionally, the ENC was also used to estimate the degree of codon usage bias of the SVA strains [20]. The value of ENC ranged from 54.51 to 55.54 with an average of 54.87 ± 0.285, which reveals a relatively stable nucleotide composition. The higher ENC of all 23 SVA strains indicate a slightly lower (ENC>40) codon usage bias.

4.3. The role of mutational pressure in shaping the codon usage bias

ENC-plot. To further investigate the pattern of the synonymous codon usage, a plot of the ENC against the GC3s was generated (Fig. 1). The results showed that all the points were clustered together and had few changes among them, indicating that the variation of the ENC was slight. It was also in accordance with the small ENC SD (0.285). Moreover, the ENC are all under the standard curve, which implies that the mutational pressure combined with other factors contributed to the codon usage bias in SVA [24].

The correlation analysis between the nucleotide contents (A%, U%, G%, C%, GC%) and the codon compositions (A3s, U3s, C3s, G3s, GC3s) showed that each had a significant correlation with the others. Moreover, the ENC correlated with the nucleotide contents with p values lower than 0.01, indicating that mutational pressure impacts the codon usage pattern of porcine SVAs.

4.4. Principal component analysis (PCA) or correspondence analysis (COA)

To investigate the codon usage pattern trends of porcine SVA, the PCA was performed, which significantly reveals the corresponding distribution of synonymous codons [15]. The PCA results displayed that the first principal axis accounted for 48.71% of all variations, which had a major impact on codon usage bias [25]. The second, the third and the fourth axes accounted for 17.6%, 10.79%, 7.64% of all variations, respectively. The first two axes are considered to be a major role in the variation of the RSCU. Therefore, a plot of the 1st and 2nd axes of the isolated strains according to the date and country of isolation were drawn (Fig. 2). It was showed that SVAs isolated from different countries were dispersed and that the strains isolated from the same countries were clustered together, except for the USA strains isolated from 2008, indicating that location could dramatically influence the codon usage bias. The correlation analysis between the codon compositions and the 1st and the 2nd axes signified the codon compositions were significantly correlated with the 1st axis (Table 4), thus confirming the mutational pressure contributed to the codon usage bias of the SVAs.

4.5. The role of natural selection in the codon usage bias of the porcine SVAs

The correlation analysis was employed to estimate the relationship between the codon usage bias and the Gravy and Aroma

Table 1
The nucleotide content of the 25 SVA strains.

Accession number	Nucleotide content			
	A%	C%	G%	U%
KX019804	23.59	28.63	23.07	24.72
KU051394	23.59	28.60	23.07	24.75
KU051393	23.59	28.70	23.05	24.66
KU051392	23.59	28.70	23.05	24.66
KU051391	23.62	28.64	23.04	24.70
KT827251	23.60	28.64	23.04	24.72
KU058183	23.57	28.69	23.07	24.67
KU058182	23.57	28.66	23.07	24.70
KU359214	23.57	28.69	23.05	24.69
KU359213	23.59	28.67	23.04	24.70
KU359212	23.59	28.66	23.04	24.72
KU359211	23.56	28.69	23.07	24.69
KU359210	23.57	28.66	23.07	24.70
KT757282	23.59	28.63	23.10	24.69
KT757281	23.57	28.61	23.08	24.73
KT757280	23.57	28.70	23.07	24.66
NC_011349	23.71	28.15	22.95	25.15
DQ641257	23.71	28.15	22.99	25.15
KR063109	23.72	28.44	22.98	24.85
KR063108	23.65	28.49	23.04	24.82
KR063107	23.65	28.49	23.04	24.82
KT321458	23.79	28.28	22.93	25.01
KC667560	23.72	28.51	22.96	24.79
Mean value ± SD	23.62 ± 0.06	28.57 ± 0.17	23.04 ± 0.05	24.77 ± 0.15

Table 2
Analysis the codon compositions (A3s%, U3s%, G3s,%,C3s%, and GC3s%).

No	A%	A3s%	C%	C3s%	G%	G3s%	U%	T3s%	GC%	GC3s%
KX019804	23.59	22.35	28.63	41.18	23.07	26.71	24.72	31.79	51.15	56.74
KU051394	23.59	22.30	28.60	41.13	23.07	26.79	24.75	31.84	51.67	56.74
KU051393	23.59	22.28	28.70	41.48	23.05	26.84	24.66	31.44	51.76	57.06
KU051392	23.59	22.28	28.70	41.48	23.05	26.84	24.66	31.44	51.76	57.06
KU051391	23.62	22.41	28.64	41.34	23.04	26.64	24.70	31.62	51.68	56.83
KT827251	23.60	22.41	28.64	41.18	23.04	26.64	24.72	31.79	51.68	56.69
KU058183	23.57	22.23	28.69	41.38	23.07	26.91	24.67	31.55	51.76	57.01
KU058182	23.57	22.24	28.66	41.38	23.07	26.92	24.70	31.55	51.73	57.01
KU359214	23.57	22.17	28.69	41.43	23.05	26.97	24.69	31.50	51.74	57.10
KU359213	23.59	22.23	28.67	41.38	23.04	26.91	24.70	31.55	51.71	57.01
KU359212	23.59	22.23	28.66	41.32	23.04	26.91	24.72	31.60	51.70	56.97
KU359211	23.56	22.11	28.69	41.43	23.07	27.04	24.69	31.50	51.76	57.15
KU359210	23.57	22.23	28.66	41.30	23.07	26.97	24.70	31.53	51.73	57.01
KT757282	23.59	22.35	28.63	41.23	23.10	26.84	24.69	31.62	51.73	56.87
KT757281	23.57	22.24	28.61	41.18	23.08	26.86	24.73	31.79	51.70	56.83
KT757222	23.57	22.22	28.70	41.48	23.07	26.91	24.66	31.44	51.77	57.10
KT321458	23.79	23.04	28.28	40.27	22.93	25.93	25.01	32.73	51.21	55.41
KR063109	23.72	23.01	28.44	40.38	22.98	26.30	24.85	32.26	51.42	55.77
KR063108	23.65	22.68	28.49	40.58	23.04	26.59	24.82	32.08	51.53	56.14
KR063107	23.65	22.68	28.49	40.58	23.04	26.59	24.82	32.08	51.53	56.14
KC667560	23.72	22.5	28.51	41.17	22.96	26.76	24.79	31.77	51.48	56.69
NC_011349	23.71	22.82	28.15	40.07	22.95	26.53	25.15	32.64	51.15	55.64
DQ641257	23.71	22.82	28.15	40.07	22.99	26.53	25.15	32.64	51.15	55.64

Table 3
The synonymous codon usage pattern in the SVA. The bolded are the preferentially used codons and RSCU values for the SVA.

AA	Codon	RSCU/Number	AA	Codon	RSCU/Number
A(Ala)	GCA	0.472/460	P(Pro)	CCA	0.407/350
	GCC	**1.667/1627**		**CCC**	**1.607/1383**
	GCG	0.579/565		CCG	0.549/473
	GCU	1.282/1251		CCU	1.437/1237
C(Cys)	TGC	0.941/217	Q(Gln)	**CAA**	**1.033/999**
	UGU	**1.059/244**		CAG	0.967/936
D(Asp)	**GAC**	**1.247/1799**	R(Arg)	**AGA**	**1.338/476**
	GAU	0.753/1086		AGG	0.523/186
E(Glu0)	GAA	0.960/1020		CGA	0.720/256
	GAG	**1.040/1104**		CGC	1.156/411
F(Phe)	**UUG**	**1.140/1547**		CGG	0.945/336
	UUU	0.860/1167		CGU	1.319/469
G(Gly)	GGA	0.956/834	S(Ser)	AGC	0.682/455
	GGC	**1.143/997**		AGU	0.610/407
	GGG	0.816/712		UCA	0.721/481
	GGU	1.085/947		UCC	1.294/863
H(His)	**CAC**	**1.121/542**		UCG	0.532/355
	CAU	0.879/425		**UCU**	**2.161/1441**
I(Ile)	AUA	0.407/250	T(Thr)	ACA	0.938/835
	AUC	**1.478/908**		**ACC**	**1.542/1373**
	AUU	1.115/685		ACG	0.443/395
K(Lys)	AAA	0.906/1037		ACU	1.077/959
	AAG	**1.094/1251**	V(Val)	GUA	0.604/539
L(Leu)	CUA	0.658/533		**GUC**	**1.306/1167**
	CUC	1.446/1173		GUG	1.225/1094
	CUG	**1.583/1283**		GUU	0.865/773
	CUU	1.059/858	Y(Tyr)	UAC	1.548/1443
	UUA	0.326/264		UAU	0.452/421
	UUG	0.928/752			
N(Asn)	**AAC**	**1.149/1398**			
	AAU	0.851/1036			

The preferentially used codons and RSCU values for the SVA are in bold.

Fig. 1. Plot of the ENC value versus the GC3s for the SVA. The dotted line is composed of the expected ENC values. The 2008 USA represented by gray, 2013 Canada represented by pink, 2015 China represented by red, 2015 Brazil represented by green, 2015 USA represented by blue, 2016 USA represented by purple. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 2. The first two axes in the correspondence analysis based on the RSCU values of SVA. The 2008 USA represented by gray, 2013 Canada represented by pink, 2015 China represented by red, 2015 Brazil represented by green, 2015 USA represented by blue, 2016 USA represented by purple. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

scores (Table 4), which demonstrated the influence of the natural selection. The results revealed that the Gravy and Aroma scores had no correlation with the codon compositions; only the Aroma score was slightly correlated with the ENC, indicating that natural selection plays role in shaping the codon usage bias of the SVAs.
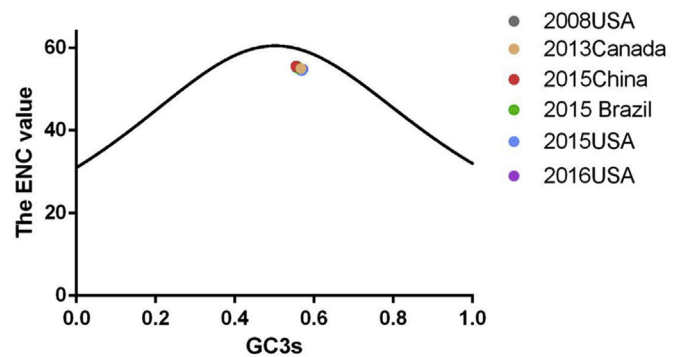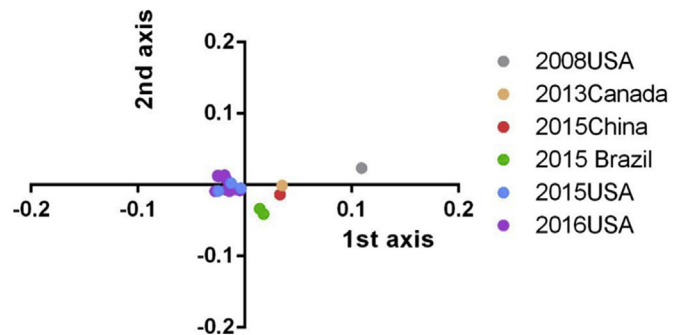
Table 4
The correlation between the codon compositions (A3s, U3s, G3s, C3s, and GC3s), the ENC values, nucleotide compositions (A%, U%, G%, C%, and GC%), the first axis values, the second axis values, the Gravy values, and the Aroma values of the SVA.

	A%	C%	G%	U%	GC%	1st axis	2nd axis	Gravy	Aroma
U3s	0.868**	0.965**	0.772**	0.950**	0.865**	0.857**	−0.137	−0.330	−0.228
C3s	0.841**	0.958**	0.729**	0.938**	0.829**	−0.864**	0.193	0.318	0.296
A3s	0.934**	0.869**	0.810**	0.821**	0.781**	0.744**	−0.384	−0.184	−0.351
G3s	0.885**	0.742**	0.765**	0.696**	0.726**	−0.571**	0.424*	0.004	0.259
ENC	0.877**	0.813**	0.797**	0.780**	0.740**	0.702**	−0.286	−0.325	−0.510*
GC3s	0.900**	0.939**	0.786**	0.910**	0.836**	−0.820**	0.261	0.255	0.294

*P<0.05, **P<0.01.

4.6. Neutrality plot analysis

The neutrality plot relationship between the GC12s and the GC3s (Fig. 3A) was employed to investigate the mutation-selection equilibrium that shapes codon usage in SVAs. The analysis showed that the GC3s was positively correlated with the GC12s (r = 0.2239, p = 0.3045). The slope of the neutrality plot was 0.03608, meaning that mutation pressure accounted for 3.608% of shaping the codon usage pattern and the influence from other factors was 96.392% [26].

Additionally, the mean GC12s and the GC3s were plotted against the isolation date (Fig. 3B). The results showed that both of the GC3s and the GC12s were positively correlated with the date (r = 0.8662, r = 0.7493), with slopes of 0.001474 ± 6.01E-04 and 0.0003158 ± 1.97E-04, respectively. This suggests that the speed of change of the GC3s is far greater than that of the GC12s, indicating that mutational pressure is increasing in shaping the SVA codon usage bias compared with the natural selection [15].

4.7. Other factors influencing the codon usage bias

In this study, the possible role of geographic distribution and time of isolated in shaping the codon usage bias of the SVAs were investigated. The average ENC of the different isolation locations and dates are shown in Fig. 4A and B, respectively, and the differences in ENC by isolation location and date were analyzed by one-way ANOVA, which showed that the mean ENC had a significant correlation with the location and date of isolation (p < 0.01). Therefore, it demonstrated that the geographic distribution and the isolation date contributed to the codon usage pattern of SVAs.
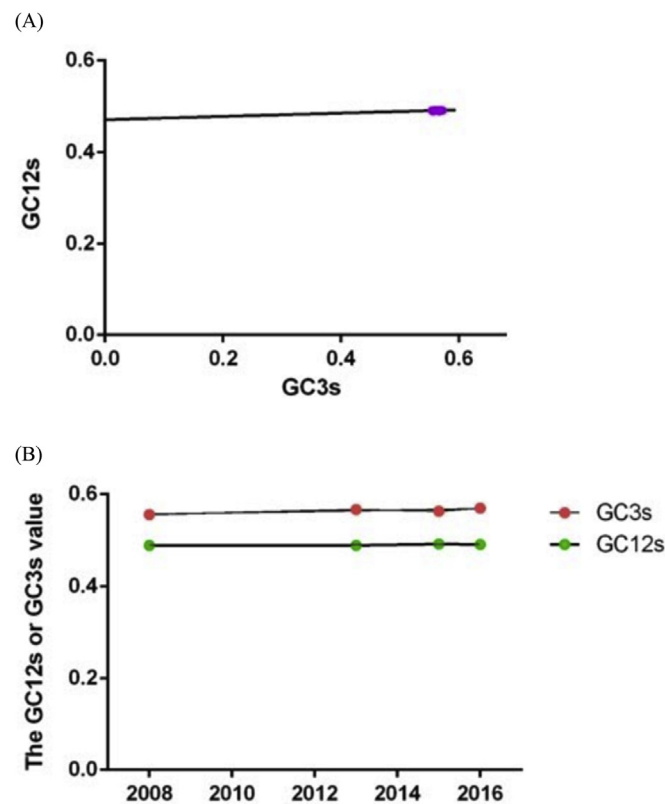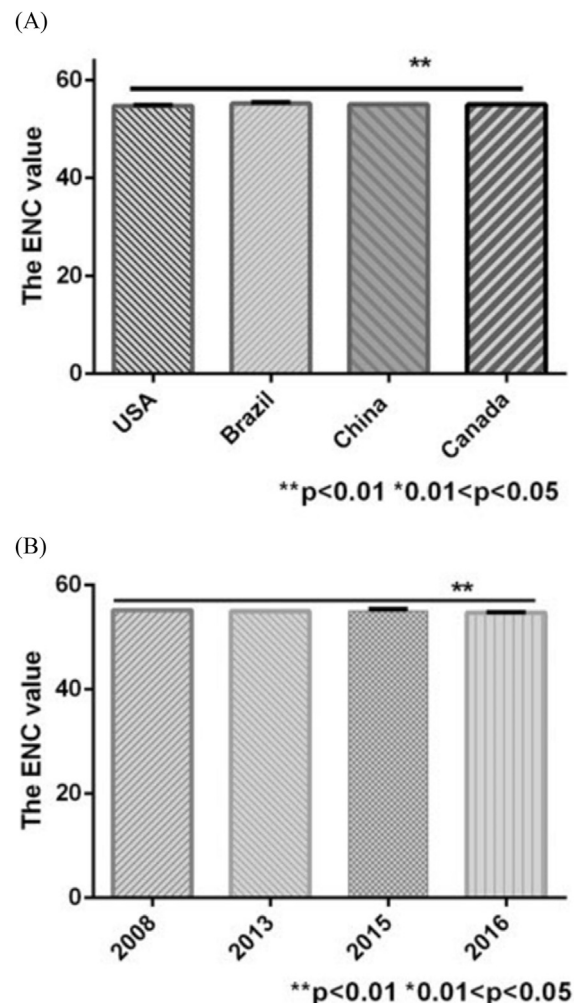

Fig. 3. (A) The neutral analysis plot of P12 against P3. (B) The evolutional analysis of the P12 and P3 values. The solid line represents the regression line.

Fig. 4. The average ENC of the different isolation locations (A) and dates (B).

5. Discussion

Compared with DNA virus, it's discovered that RNA virus, such as influenza virus, coronaviruses, exists a highly evolution rate [27–30]. Previous studies revealed that the prototype SVA isolate, namely SVV-001, was isolated in 2001 [31], and due to its selective tropism for human tumor cells, the virus was developed as an oncolytic agent [32].However, the basic viral pathogenesis was not detected accurately. In this study, the synonymous codon usage of the SVA was analyzed to better understand the SVA evolution, especially the interplay between the viruses and the immune response [33]. In the analysis, the ENC and RSCU were calculated. The ENC ranged from 54.51 to 55.54 with a mean ENC 54.87, which indicates a low codon usage bias. This result is in accordance with other *Picornaviridae* viruses, such as poliovirus type A9 (ENC = 54.2), foot-and-mouth disease virus (mean ENC = 51.53), rhinovirus type 89 (ENC = 45.9), and enterovirue 71 (ENC = 56.6) [34]. The results of overall RSCU of all the codons which revealed that most of the preferentially used codons are G/C-ended codons, combined with the mean GC3s (56.6%) suggested that (G + C) compositions constrain the codon usage bias of the SAV. This discovery is in agreement with the previous reports that G + C contents are the dominant factors influencing the codon usage in virus genomes [35,36].

Although the codon usage bias of SVA is low, the factors exist in it were analyzed. Normally, mutational pressure and natural selection are considered to be the main factors influencing the codon usage variation among genes. In this study, as shown in Fig. 1, all the strains are clustered together, and the observed ENC are just below the standard curve, revealing that mutational pressure affects the codon usage bias. Additionally, the PCA suggested that the different isolated strains were diverged from each other and did not locate on or around the origin of coordinate, which also explained that there exists mutational pressure in shaping the codon usage pattern. Furthermore, the results were confirmed by the correlation analysis between the nucleotide content (A%, TU%, G%, C%, GC%) and the codon compositions (A3s, U3s, C3s, G3s, GC3s), which each other had significant differences and existed positively correlation. For natural selection, only the Aroma score was slightly correlated with the ENC, and other indicies had no correlation, demonstrated that natural selection plays a role in the codon usage bias of SVS.

Actually, the mutational pressure is considered as the major factor in shaping the codon usage variation compared with natural selection [37]. Thus, to determine which was responsible for the extreme codon usage bias in SVA, the GC12s and GC3s were plotted against the evolution time in an evolutional neutrality analysis, and the results revealed that the mutational pressure plays an increasing role in the SVA evolution compared with natural selection. In addition, the neutrality plot showed that the relationship between GC3s and GC12s was not significant, and the correlation coefficient indicated that compared with mutation pressure, natural selection is more important in codon usage bias of SVA. As shown in Fig. 2, the 23 strains isolated from different location or dates were separated, suggesting that the geographic distribution and date of isolation may contribute to the codon usage pattern of SVA. To confirm this hypothesis, the relationship between the date or location of isolation and the ENC were analyzed, respectively. The significant difference among ENC value and isolated country and year indicated that the geographical distribution and isolated year are essential to some extent.

6. Conclusions

In general, this study first identified a low codon usage bias in the SVA codon usage pattern and that the mutational pressure is an increasingly dominant factor in the evolution of this virus compared with natural selection. Moreover, codon usage bias was affected by the geographic distribution and date of isolation. The systemic analysis of the codon usage bias of the SVA is beneficial to understand the evolution of SVA, which provides scientific proof for the directorial department to make preventive measures against SVA.

Declarations

Ethics

Not applicable.

Consent to publish

Not applicable.

Availability of data and material

The datasets and materials used and/or analyzed during the current study available from the Genebank by the corresponding author on reasonable request.

Abbreviations

SVA   Senecavirus A
SVV   Seneca Valley virus
SVV-001   Seneca Valley virus 001
PEDV   porcine epidemic diarrhea virus
NCBI   National Center for Biotechnology
GC1s, GC2s, GC3s G + C at the first, second, third
RSCU   The relative synonymous codon usage
ENC   Effective number of codons
PCA   Principal component analysis
A3, U3, G3, C3, GC3 the nucleotide composition at the third codon
                  position
COA   Correspondence analysis

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.micpath.2017.09.040.

References

[1] R.A. Leme, T.E. Oliveira, B.K. Alcantara, S.A. Headley, A.F. Alfieri, M. Yang, A.A. Alfieri, Clinical manifestations of Senecavirus A infection in neonatal pigs, Brazil, 2015, Emerg. Infect. Dis. 22 (7) (2016) 1238–1241.
[2] L.M. Hales, N.J. Knowles, P.S. Reddy, L. Xu, C. Hay, P.L. Hallenbeck, Complete genome sequence analysis of Seneca Valley virus-001, a novel oncolytic picornavirus, J. Gen. Virol. 89 (Pt 5) (2008) 1265–1275.
[3] L.M. Hales, B.H. Jones, A.-J. Vasko, N.J. Knowles, S.R. Police, P.L. Hallenbeck, EPIDEMIOLOGY of Seneca Valley virus (SVV-001), a novel oncolytic picornavirus for the systemic treatment of patients with solid cancers with neuroendocrine features, Mol. Ther. 13 (S1) (2006) S187.
[4] N.J. Knowles, L.M. Hales, B.H. Jones, J.G. Landgraf, J.A. House, K.L. Skele, K.D. Burroughs, P.L. Hallenbeck, Epidemiology of Seneca Valley Virus: Identification and Characterization of Isolates from Pigs in the United States, 2006. Northern Lights EUROPIC.
[5] S. Venkataraman, S.P. Reddy, J. Loo, N. Idamakanti, P.L. Hallenbeck, V.S. Reddy, Structure of Seneca Valley Virus-001: an oncolytic picornavirus representing a new genus, Structure 16 (10) (2008) 1555–1561.

[6] H.L. Bachrach, D.M. Moore, P.D. McKercher, J. Polatnick, Immune and antibody responses to an isolated capsid protein of foot-and-mouth disease virus, J. Immunol. 115 (6) (1975) 1636—1641.

[7] E. Ehrenfeld, E. Domingo, R.P. Roos, The Picornaviruses, American Society for Microbiology Press, 2010.

[8] F.A. Vannucci, D.C. Linhares, D.E. Barcellos, H.C. Lam, J. Collins, D. Marthaler, Identification and complete genome of Seneca Valley virus in vesicular fluid and sera of pigs affected with idiopathic vesicular disease, Brazil, Transbound. Emerg. Dis. 62 (6) (2015) 589—593.

[9] R.A. Leme, E. Zotti, B.K. Alcantara, M.V. Oliveira, L.A. Freitas, A.F. Alfieri, A.A. Alfieri, Senecavirus A: an emerging vesicular infection in brazilian pig herds, Transbound. Emerg. Dis. 62 (6) (2015) 603—611.

[10] B.R.C. Guo, Y. Zheng, Yoon Identification of novel Senecavirus A from pigs with vesicular disease in the US, in: 24th International Pig Veterinary Society Congress (IPVS), 2016, p. 109.

[11] J. Segales, D. Barcellos, A. Alfieri, E. Burrough, D. Marthaler, Senecavirus A: an emerging pathogen causing vesicular disease and mortality in pigs? Veterinary pathol. 54 (1) (2016) 11—21.

[12] P. Canning, A. Canon, J.L. Bates, K. Gerardy, D.C. Linhares, P.E. Pineyro, K.J. Schwartz, K.J. Yoon, C.J. Rademacher, D. Holtkamp, et al., Neonatal mortality, vesicular lesions and lameness associated with Senecavirus A in a U.S. Sow farm, Transbound. Emerg. Dis. 63 (4) (2016) 373—378.

[13] Q. Wu, X. Zhao, Y. Bai, B. Sun, Q. Xie, J. Ma, The first identification and complete genome of Senecavirus A affecting pig with idiopathic vesicular disease in China, Transbound. Emerg. Dis. (2016).

[14] A. Marin, J. Bertranpetit, J.L. Oliver, J.R. Medina, Variation in G + C-content and codon choice: differences among synonymous codon groups in vertebrate genes, Nucleic acids Res. 17 (15) (1989) 6181—6189.

[15] Y. Chen, Y. Shi, H. Deng, T. Gu, J. Xu, J. Ou, Z. Jiang, Y. Jiao, T. Zou, C. Wang, Characterization of the porcine epidemic diarrhea virus codon usage bias, Infect. Genet. Evol. 28 (2014) 95—100.

[16] Y. Chen, X. Sun, X. Tong, J. Xu, H. Deng, Z. Jiang, C. Jiang, J. Duan, J. Li, P. Zhou, et al., First analysis of synonymous codon usage in porcine circovirus, Archives Virol. 159 (8) (2014) 2145—2151.

[17] J. Cristina, A. Fajardo, M. Sonora, G. Moratorio, H. Musto, A detailed comparative analysis of codon usage bias in Zika virus, Virus Res. 223 (2016) 147—152.

[18] P.M. Sharp, W.H. Li, Codon usage in regulatory genes in Escherichia coli does not reflect selection for 'rare' codons, Nucleic acids Res. 14 (19) (1986) 7737—7749.

[19] J.H. Zhou, Y.Z. Ding, Y. He, Y.F. Chu, P. Zhao, L.Y. Ma, X.J. Wang, X.R. Li, Y.S. Liu, The effect of multiple evolutionary selections on synonymous codon usage of genes in the Mycoplasma bovis genome, PLoS One 9 (10) (2014) e108949.

[20] J.M. Comeron, M. Aguade, An evaluation of measures of synonymous codon usage bias, J. Mol. Evol. 47 (3) (1998) 268—274.

[21] F. Wright, The 'effective number of codons' used in a gene, Gene 87 (1) (1990) 23—29.

[22] N. Kumar, B.C. Bera, B.D. Greenbaum, S. Bhatia, R. Sood, P. Selvaraj, T. Anand, B.N. Tripathi, N. Virmani, Revelation of influencing factors in overall codon usage bias of equine influenza viruses, PLoS One 11 (4) (2016) e0154376.

[23] S.K. Gupta, T.C. Ghosh, Gene expressivity is the main factor in dictating the codon usage variation among the genes in Pseudomonas aeruginosa, Gene 273 (1) (2001) 63—70.

[24] N. Sueoka, Directional mutation pressure and neutral molecular evolution, Proc. Natl. Acad. Sci. U. S. A. 85 (8) (1988) 2653—2657.

[25] L. Peixoto, A. Zavala, H. Romero, H. Musto, The strength of translational selection for codon usage varies in the three replicons of Sinorhizobium meliloti, Gene 320 (1) (2003) 109—116.

[26] X. Li, H. Song, Y. Kuang, S. Chen, P. Tian, C. Li, Z. Nan, Genome-wide analysis of codon usage bias in epichloe festucae, Int. J. Mol. Sci. 17 (7) (2016).

[27] S. Su, Y. Bi, G. Wong, G.C. Gray, G.F. Gao, S. Li, The epidemiology, evolution and recent outbreaks of avian influenza viruses in China: a review, J. Virol. 89 (17) (2015). JVI.01034-01015.

[28] S. Su, G. Wong, W. Shi, J. Liu, A.C. Lai, J. Zhou, W. Liu, Y. Bi, G.F. Gao, Epidemiology, genetic recombination, and pathogenesis of coronaviruses, Trends Microbiol. 24 (6) (2016) 490.

[29] G. Li, S. Ji, X. Zhai, Y. Zhang, L. Jie, M. Zhu, J. Zhou, S. Su, Evolutionary and genetic analysis of the VP2 gene of canine parvovirus, Bmc Genomics 18 (1) (2017) 534.

[30] S. Su, M. Gu, D. Liu, J. Cui, G.F. Gao, J. Zhou, X. Liu, Epidemiology, evolution, and pathogenesis of H7N9 influenza viruses in five epidemic waves since 2013 in China, Trends Microbiol. 25 (9) (2017) 713.

[31] A.J. Bracht, E.S. O'Hearn, A.W. Fabian, R.W. Barrette, A. Sayed, Real-time reverse transcription PCR assay for detection of Senecavirus A in swine vesicular diagnostic specimens, PLoS One 11 (1) (2016) e0146211.

[32] Z. Chen, F. Yuan, Y. Li, P. Shang, R. Schroeder, K. Lechtenberg, J. Henningson, B. Hause, J. Bai, R.R. Rowland, et al., Construction and characterization of a full-length cDNA infectious clone of emerging porcine Senecavirus A, Virology 497 (2016) 111—124.

[33] L.A. Shackelton, C.R. Parrish, E.C. Holmes, Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses, J. Mol. Evol. 62 (5) (2006) 551—563.

[34] Y. Zhang, Y. Liu, W. Liu, J. Zhou, H. Chen, Y. Wang, L. Ma, Y. Ding, J. Zhang, Analysis of synonymous codon usage in hepatitis A virus, Virol. J. 8 (2011) 174.

[35] S. Zhao, Q. Zhang, X. Liu, X. Wang, H. Zhang, Y. Wu, F. Jiang, Analysis of synonymous codon usage in 11 human bocavirus isolates, Bio Syst. 92 (3) (2008) 207—214.

[36] S. Das, S. Paul, C. Dutta, Synonymous codon usage in adenoviruses: influence of mutation, selection and protein hydropathy, Virus Res. 117 (2) (2006) 227—236.

[37] G.M. Jenkins, E.C. Holmes, The extent of codon usage bias in human RNA viruses and its evolutionary origin, Virus Res. 92 (1) (2003) 1—7.