Microbial Pathogenesis 112 (2017) 313–319

Contents lists available at ScienceDirect

Microbial Pathogenesis

journal homepage: www.elsevier.com/locate/micpath

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

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

Article history:
Received 22 August 2017
Received in revised form
16 September 2017
Accepted 18 September 2017
Available online 22 September 2017

Keywords:
Senecavirus A
Codon usage
Mutation pressure
Natural selection

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.

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.

https://doi.org/10.1016/j.micpath.2017.09.040
0882-4010/

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:

$$\mathrm{RSCU} = \frac{g_{ij}}{\sum_{j}^{ni} g_{ij}} ni$$

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:

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

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:

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

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