Virus Research 223 (2016) 73–79

Contents lists available at ScienceDirect

Virus Research

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

Short communication

Analysis of the synonymous codon usage bias in recently emerged enterovirus D68 strains

Uladzimir U. Karniychuk

Vaccine and Infectious Disease Organization-International Vaccine Centre (VIDO-InterVac), University of Saskatchewan, Saskatoon, SK S7N 5E3, Canada

ARTICLE INFO

Article history:
Received 5 February 2016
Received in revised form 9 May 2016
Accepted 24 June 2016
Available online 27 June 2016

Keywords:
Enterovirus D68
Synonymous codon usage bias
Animal models

ABSTRACT

Understanding the codon usage pattern of a pathogen and relationship between pathogen and host's codon usage patterns has fundamental and applied interests. Enterovirus D68 (EV-D68) is an emerging pathogen with a potentially high public health significance. In the present study, the synonymous codon usage bias of 27 recently emerged, and historical EV-D68 strains was analyzed. In contrast to previously studied enteroviruses (enterovirus 71 and poliovirus), EV-D68 and human host have a high discrepancy between favored codons. Analysis of viral synonymous codon usage bias metrics, viral nucleotide/dinucleotide compositional parameters, and viral protein properties showed that mutational pressure is more involved in shaping the synonymous codon usage bias of EV-D68 than translation selection. Computation of codon adaptation indices allowed to estimate expression potential of the EV-D68 genome in several commonly used laboratory animals. This approach requires experimental validation and may provide an auxiliary tool for the rational selection of laboratory animals to model emerging viral diseases. Enterovirus D68 genome compositional and codon usage data can be useful for further pathogenesis, animal model, and vaccine design studies.

The degeneracy of the genetic code, or a synonymous codon usage, implies the usage of more than one codon to encode for the same amino acid. The synonymous codon usage is biased; different synonymous codons in a variety of evolutionarily distant species are overrepresented or underrepresented (Grantham et al., 1980; Lipman and Wilbur, 1983; Shackelton et al., 2006; Sharp and Li, 1986). A number of forces, like mutational pressure, natural selection (such as translational selection) (Jenkins and Holmes, 2003; Sharp and Li, 1986; Sharp et al., 2010), the control of the speed of translation as a way to control protein folding, and the need to escape antiviral host responses (Pintó et al., 2012, 2007), can affect this synonymous codon usage bias (SCUB). Detailed understanding of the nature and cause of SCUB has a general evolutionary (Grantham et al., 1980; Shackelton et al., 2006) and applied interests (Burns et al., 2006; Mueller et al., 2006). The present computational biology research was focused on the role of mutational pressure and translational selection in shaping enterovirus D68 (EV-D68) SCUB.

Enterovirus D68 is an emerging pathogen with a potentially high public health significance (Imamura and Oshitani, 2015; Tokarz et al., 2012). The virus is a member of species enterovirus D, genus Enterovirus, and family Picornaviridae. Enterovirus D68 is a non-enveloped, single-stranded, positive-sense RNA virus. A viral genome contains a single open reading frame encoding four structural (VP1–VP4) and seven non-structural proteins (2A–2C and 3A–3D) (Imamura and Oshitani, 2015). Enterovirus D68 infection causes acute respiratory disorders mostly in children. A considerable number of cases is severe, and some are fatal (Foster et al., 2015; Imamura and Oshitani, 2015; Meijer et al., 2014; Tokarz et al., 2012). The most recent outbreak of severe pneumonia in children caused by EV-D68 has been reported in the USA and Canada in the fall of 2014 (Drews et al., 2015; Foster et al., 2015). In addition to the respiratory disease, EV-D68 infection is also associated with neurological disorders (Messacar et al., 2015). In contrast to other human picornaviruses like hepatitis A virus (HAV), poliovirus and enterovirus 71 (EV-71) (Zhang et al., 2014; D'Andrea et al., 2011; Liu et al., 2011; Ma et al., 2014; Zhang et al., 2011), there are no studies on EV-D68 SCUB and on an importance of mutational pressure and translational selection in shaping EV-D68 SCUB. In the present study, I analyzed SCUB of historical and recently emerged EV-D68 strains and compared EV-D68 SCUB with other picornaviruses and human SCUB. Several metrics reflecting mutational pressure and translational selection, and potential involvement of these evolutionary forces in shaping EV-D68 SCUB were analyzed as well. Also, the expression potential of EV-D68 coding sequences (CDSs)

E-mail address: u.karniychuk@usask.ca

http://dx.doi.org/10.1016/j.virusres.2016.06.015
0168-1702/

in the natural host and several laboratory animals commonly used in research were theoretically assessed.

All viral complete CDSs were downloaded from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/Genbank/). Sequences with more than 99% identity were excluded, and dataset represented 27 EV-D68 CDSs (STable 1). Two commonly used metrics reflecting SCUB are relative synonymous codon usage (RSCU) and observed effective number of codons (ENC). The relative synonymous codon usage is the ratio of the observed frequency of a codon to its expected frequency assuming the equal use of all synonymous codons (Shackelton et al., 2006; Sharp and Li, 1986). Because its independence from an encoding amino acid composition, RSCU is useful for SCUB comparison between genes with different size and amino acid composition, allowing to contrast codon usage preferences between virus and host. To compute RSCU, EV-D68 CDSs were processed with the SSE_v1.2 editor (Simmonds, 2012). Table 1 represents mean RSCU values and codon usage preference of 27 EV-D68 CDSs and human genes (Cristina et al., 2015; D'Andrea et al., 2011). Enterovirus D68 has the A and U-rich (A + U content is 59%), and G and C-poor (G + C content is 41%) genome (STable 1). In accordance with these compositional constraints, the preferentially used EV-D68 codons are A-ended (10 codons) and U-ended (7 codons), and only Tyr is preferentially coded by the triplet ended with C (Table 1). The relative synonymous codon usage in EV-D68 CDSs are considerably different from RSCU of the host (Table 1). Only 2 of 59 codons have the usage preference in both EV-D68 and human cells, including UAC for Tyr and AGA for Arg. Particularly highly biased frequencies (EV-D68 RSCU – human RSCU $\geq$ 0.30) were found for Leu, Ile, Val, Ser, Pro, Thr, Ala, His, Asn, Lys, Asp, Glu, Cys, Arg and Gly (Table 1). Among previously analyzed picornaviruses, hepatitis A virus (HAV) has similar biased codon usage preferences in comparison to the host (1 of 59 codons has the usage preference in both HAV and human cells) (D'Andrea et al., 2011; Pintó et al., 2007; Sánchez et al., 2003). In contrast, RSCU is in the higher agreement between poliovirus (7 of 59 codons have the usage preference in both poliovirus and human cells), enterovirus 71 (EV-71) (8–10 of 59 codons have the usage preference in both EV-71 and human cells) and host (Zhang et al., 2014; Liu et al., 2011; Ma et al., 2014; Zhang et al., 2011). Like HAV (Aragonès et al., 2010; Costafreda et al., 2014), EV-D68 might synthesize proteins adapting the codon usage to less commonly used human transfer RNA (tRNAs), avoiding competition with host genes for most abundant tRNAs. To support this hypothesis, the frequencies of tRNAs in human cells were retrieved from the GtRNAdb database (Chan and Lowe, 2009), and analysis revealed that in most cases (ACU, GAA, and AGA codons are exceptions), EV-D68 uses the less abundant human tRNAs (STable 2).

In contrast to RSCU, ENC reflects the absolute synonymous codon usage (Wright, 1990). The observed effective number of codons in EV-D68 CDSs were computed with the SSE_v1.2 editor (Simmonds, 2012). The ENC values range from 20 to 61 (Wright, 1990). The bigger the level of SCUB in a gene, the smaller the ENC value. In an extremely biased gene where only one codon is used for each amino acid, ENC is 20. In an unbiased gene where codons are used equally, ENC is 61. The values of ENC among 27 EV-D68 CDSs are similar, ranging from 46.25 to 48.79, with a mean 48.09 (standard deviation (S.D.) 0.6) (STable 1). Comparing with other human picornaviruses, HAV has higher SCUB (mean 39.34–39.78) (D'Andrea et al., 2011) than EV-D68, while poliovirus (53.72–53.75) (Zhang et al., 2011) and EV-71 (56.62) (Liu et al., 2011; Ma et al., 2014; Zhang et al., 2014) have lower SCUB than EV-D68. A plot of ENC values as a function of the GC3 (G + C frequency at the third codon position) content provides a useful visualization of the main synonymous codon usage trends (Wright, 1990). Observed ENC and expected ENC values were plotted versus GC3 frequencies in

Table 1
Synonymous codon usage bias in EV-D68 coding sequences.

Amino acid	Codon	RSCU	
		EV-D68[a]	HS[b]
Phe	UUU	**1.18**	0.92
	UUC	0.82	**1.08**
Leu	UUA	**1.82**	0.48
	UUG	1.05	0.78
	CUU	1.04	0.78
	CUC	0.51	1.2
	CUA	1.01	0.42
	CUG	0.56	**2.4**
Ile	AUU	**1.15**	1.08
	AUC	0.65	**1.41**
	AUA	1.09	0.51
Met	AUG	1	1
Val	GUU	**1.18**	0.72
	GUC	0.68	0.96
	GUA	1.06	0.48
	GUG	1.07	**1.84**
Ser	AGU	1.53	0.9
	AGC	0.64	**1.44**
	UCU	1.09	1.14
	UCC	0.5	1.32
	UCA	**2.02**	0.9
	UCG	0.22	0.3
Pro	CCU	0.79	1.16
	CCC	0.65	**1.28**
	CCA	**2.32**	1.12
	CCG	0.23	0.44
Thr	ACU	1.39	1
	ACC	0.84	**1.44**
	ACA	**1.59**	1.12
	ACG	0.18	0.44
Ala	GCU	1.27	1.08
	GCC	0.8	**1.6**
	GCA	**1.72**	0.92
	GCG	0.21	0.44
Tyr	UAU	0.97	0.88
	UAC	**1.03**	**1.12**
TER	UAA		
	UAG		
His	CAU	**1.25**	0.84
	CAC	0.75	**1.16**
Gln	CAA	**1.16**	0.54
	CAG	0.84	**1.46**
Asn	AAU	**1.28**	0.94
	AAC	0.72	**1.06**
Lys	AAA	**1.34**	0.86
	AAG	0.66	**1.14**
Asp	GAU	**1.32**	0.92
	GAC	0.68	**1.08**
Glu	GAA	**1.28**	0.84
	GAG	0.72	**1.16**
Cys	UGU	**1.23**	0.92
	UGC	0.77	**1.08**
TER	UGA		
Trp	UGG	1	1
Arg	AGA	**3.73**	**1.26**
	AGG	1.39	**1.26**
	CGU	0.23	0.48
	CGC	0.42	1.08
	CGA	0.17	0.66
	CGG	0.06	1.2
Gly	GGU	1.34	0.64
	GGC	0.66	**1.36**
	GGA	**1.36**	1
	GGG	0.64	1

EV-D68 – Enterovirus D68.
RSCU – relative synonymous codon usage.
Termination codons (TER) and single codons encoding methionine (Met) and tryptophan (Trp) were exclude from correspondence analysis.
[a] Mean relative synonymous codon usage values of 27 EV-D68 strains.
[b] RSCU of Homo sapiens (Cristina et al., 2015; D'Andrea et al., 2011). RSCU of the preferentially used codons is in bold.