Virus Research 196 (2015) 87–93

Contents lists available at ScienceDirect

Virus Research

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

Genome-wide analysis of codon usage bias in Ebolavirus

Juan Cristina[a,*], Pilar Moreno[a], Gonzalo Moratorio[a,b], Héctor Musto[c]

[a] Laboratorio de Virología Molecular, Centro de Investigaciones Nucleares, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400 Montevideo, Uruguay
[b] Viral Populations and Pathogenesis Laboratory, Institut Pasteur, CNRS UMR 3569, Paris, France
[c] Laboratorio de Organización y Evolución del Genoma, Instituto de Biología, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400 Montevideo, Uruguay

ARTICLE INFO

Article history:
Received 25 September 2014
Received in revised form 31 October 2014
Accepted 6 November 2014
Available online 14 November 2014

Keywords:
Ebola
Codon usage
Codon bias
Evolution

ABSTRACT

Ebola virus (EBOV) is a member of the family *Filoviridae* and its genome consists of a 19-kb, single-stranded, negative sense RNA. EBOV is subdivided into five distinct species with different pathogenicities, being *Zaire ebolavirus* (ZEBOV) the most lethal species. The interplay of codon usage among viruses and their hosts is expected to affect overall viral survival, fitness, evasion from host's immune system and evolution. In the present study, we performed comprehensive analyses of codon usage and composition of ZEBOV. Effective number of codons (ENC) indicates that the overall codon usage among ZEBOV strains is slightly biased. Different codon preferences in ZEBOV genes in relation to codon usage of human genes were found. Highly preferred codons are all A-ending triplets, which strongly suggests that mutational bias is a main force shaping codon usage in ZEBOV. Dinucleotide composition also plays a role in the overall pattern of ZEBOV codon usage. ZEBOV does not seem to use the most abundant tRNAs present in the human cells for most of their preferred codons.

1. Introduction

Ebola virus (EBOV) is a member of the family *Filoviridae* and is among the most deadly human pathogens, causing a severe hemorrhagic fever syndrome in both humans and non-human primates (Hoenen et al., 2006; Sanchez et al., 2007). EBOV is subdivided into five distinct species with different pathogenicity (Wauquier et al., 2010): *Zaire ebolavirus* (ZEBOV), the most lethal species with a case-fatality rate up to 90% (Khan et al., 1999), that caused numerous human outbreaks in Democratic Republic of Congo; *Sudan ebolavirus* (SEBOV), with a case-fatality rate of about 50%, which has caused outbreaks in Sudan and Uganda (WHO, 2004; CDC, 2001); *Cote d'Ivoire ebolavirus* (CIEBOV), who has been linked to a single non-fatal human case (Le Guenno et al., 1999), the newly discovered *Bundibugyo ebolavirus* (BEBOV) that caused an outbreak with a 25% case-fatality rate in 2007 in Uganda (Towner et al., 2008), and *Reston ebolavirus* (REBOV), which has caused outbreaks in non-human primates and swine in the Philippines and appears to be non-pathogenic for humans (Rollin et al., 1999).

EBOV is a filamentous enveloped virus containing a negative strand RNA genome of approximately 19 kb. The EBOV genome consists of eight major subgenomic mRNAs which in turn encode for seven structural proteins, namely a nucleoprotein (NP), two virion proteins (VP35 and VP40), a surface glycoprotein (GP), two additional viral proteins (VP30 and VP24), and a RNA-dependent RNA polymerase (L) and for one non-structural soluble protein (sGP) (Feldmann et al., 1993).

The disease caused by EBOV is characterized by the sudden onset of fever and malaise, accompanied by other nonspecific signs and symptoms such as myalgia, headache, vomiting, and diarrhea. EBOV initially replicates massively in macrophages and dendritic cells (DC), then spreads rapidly to all vital organs, infecting endothelial cells, epithelial cells, hepatocytes and other cell types (Stroher et al., 2001; Geisbert et al., 2003). Among EBOV patients, 30–50% experience hemorrhagic symptoms. A recent report by Rasmussen et al. (2014), suggested that the genetic factors play a significant role in determining hemorrhagic outcome in naïve individuals without prior exposure or immunologic priming by using mice model. Thence, in severe and fatal forms, multiorgan dysfunctions, including hepatic damage, renal failure, and central nervous system involvement occur, leading to shock and death (Dixon and Schafer, 2014). There is currently no vaccine and no specific treatment against EBOV (Becquart et al., 2014).

* Corresponding author. Tel.: +598 2 525 09 01; fax: +598 2 525 08 95.
E-mail address: cristina@cin.edu.uy (J. Cristina).

http://dx.doi.org/10.1016/j.virusres.2014.11.005
0168-1702/

Since 1976, there have been more than 20 EBOV outbreaks across Central Africa, with the majority caused by ZEBOV. No previous EBOV outbreak has been as large or persistent as the current epidemic, and none has spread beyond East and Central Africa (Dixon and Schafer, 2014; Gire et al., 2014). To date, more than 4555 people, including health care workers, have been killed by EBOV disease in 2014 (as of October 22, 2014), and the number of cases in the current outbreak now exceeds the number from all previous outbreaks combined (Frieden et al., 2014). Viral sequencing of specimens from the current epidemic were identified as ZEBOV (Baize et al., 2014).

The redundancy of the genetic code, in which most of the amino acids can be translated by more than one codon, offers evolution the opportunity to tune the efficiency and accuracy of protein production to various levels while maintaining the same amino acid sequence (Stoletzki and Eyre-Walker, 2007). The various codons that correspond to the same residue are often considered 'synonymous', yet their corresponding tRNAs might differ in their amounts in cells and thus also in the speed in which they will be recognized by the ribosome, and hence, influence the rate of translation and the accuracy in folding the encoded protein. While the non-random usage of synonymous codons is often correctly assumed to reflect the action of neutral drift, in an increasing number of cases it now turns out to reflect the result of natural selection, perhaps mainly for tuning the efficiency and accuracy of translation (Gingold and Pilpel, 2011). Studies on codon usage have determined several factors that could influence codon usage patterns, including mutational pressure, natural selection, secondary protein structure, replication and selective transcription among others (Butt et al., 2014).

The interplay of codon usage among viruses and their hosts is expected to affect overall viral survival, fitness, evasion from host's immune system and evolution (Burns et al., 2006; Mueller et al., 2006; Costafreda et al., 2014). Indeed, as is well known, synonymous triplets are generally not used randomly, and the main forces that drive this bias from equal usage are natural selection (which is mainly related to translation efficiency at two different levels: speed and accuracy) and mutational biases (for a review see Sharp et al., 2010). Therefore, the study of codon usage patterns in viruses can reveal important information about molecular evolution, regulation of viral gene expression and aid in vaccine design, where the efficient expression of viral proteins may be required to generate immunity (Butt et al., 2014).

In the present study, we performed comprehensive analyses of codon usage and composition of ZEBOV, including the recently isolated strains from the current epidemic of 2014, which represents all the complete genome sequences available in the databases, and investigated the possible key evolutionary determinants of the biases found.

2. Materials and methods

2.1. Sequences

Complete genome sequences for 25 Zaire Ebolavirus strains (ZEBOV) were obtained from DDBJ and GeneBank databases (available at: http://arsa.ddbj.nig.ac.jp and http://www.ncbi.nlm.nih.gov, respectively). For strain names and accession numbers see Supplementary Table S1. For each strain the ORFs were concatenated (NP+VP35+VP40+GP+VP30+VP24+L) and aligned using the MUSCLE program (Edgar, 2004). The alignment is available upon request. The data set comprised a total of 113,025 codons.

Supplementary Table S1 related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.virusres.2014.11.005.

2.2. Data analysis

Codon usage, dinucleotide frequencies, base composition, the relative synonymous codon usage (RSCU) (Sharp and Li, 1986), the effective number of codons (ENC) (Novembre, 2002), total G+C genomic content, as well as G+C content at first, second and third codon positions were calculated using the program CodonW (written by John Peden and available at http://sourceforge.net/projects/codonw). ENC index can vary between 20 and 61 and a low value indicates a strong bias in codon usage. To study codon usage preferences in EBOV in relation to the codon usage of human cells, we employed the codon adaptation index (CAI) (Sharp and Li, 1987). CAI was calculated using the approach of Puigbo et al. (2008a) (available at: http://genomes.urv.es/CAIcal) for EBOV and human cells. This method allows to compare a given codon usage (in our case, EBOV) to a predefined reference set (human). In order to show whether the EBOV genes are not better adapted to the codon usage of the reference set than the genes that define the reference dataset itself, as measured by CAI, we constructed a dataset composed of 322 human genes selected at random from Ensembl database (available at http://www.ensembl.org). A statistically significant difference among CAI values obtained was addressed by means of the use of a Student's *t*-test and a Wilcoxom & Mann–Whitney test (Wessa, 2012). In order to discern if the statistically significant differences in the CAI values arise from codon preferences, we used E-CAI (Puigbo et al., 2008b) to calculate the expected value of CAI (eCAI) at the 95% confident interval. A Kolmogorov–Smirnov test for the expected CAI was also performed (Puigbo et al., 2008b). The RSCU values of human cells were obtained from Kazusa database (available at: http://www.kazusa.or.jp/codon/). The frequencies of tRNAs in human cells were retrieved from the GtRNAdb database (Chan and Lowe, 2009).

2.3. Correspondence analysis

The relationship between variables and samples can be obtained using multivariate statistical analysis. Correspondence analysis (COA) is a type of multivariate analysis that allows a geometrical representation of the sets of rows and columns in a dataset (Wong et al., 2010; Greenacre, 1984). Each ORF is represented as a 59-dimensional vector and each dimension correspond to the RSCU value of one codon (excluding AUG, UGG and stop codons). Major trends within a dataset can be determined using measures of relative inertia and genes ordered according to their position along the different axes (Tao et al., 2009). COA was performed on the RSCU values using the CodonW program.

2.4. Correlation analysis

Correlation analysis was carried out using Spearman's rank correlation analysis method (Wessa, 2012; available at: www.wessa.net).

3. Results and discussion

3.1. General codon usage pattern in ZEBOV

In order to study the extent of codon usage bias in Zaire Ebolavirus strains (ZEBOV), the ENC's values were calculated for the 25 strains enrolled in this study. A mean value of 57.23 ± 0.51 was obtained. Due to the fact that all values obtained were >40, the results of these studies suggest that the overall codon usage among ZEBOV is similar and slightly biased.

This is in agreement with previous studies in other RNA viruses, like Chikungunya virus (ENC = 55.56) (Butt et al., 2014), bovine

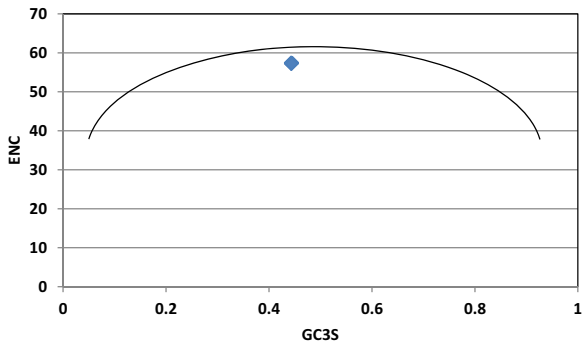

Fig. 1. Effective number of codons (ENC) used in ZEBOV ORFs plotted against the $GC3_S$. The curve plots the relationship between $GC3_S$ and ENC in absence of selection. Square dots show the results obtained for ZEBOV strains.

viral diarrheavirus (ENC = 50.91) (Wang et al., 2011), classical swine fever virus (ENC = 51.7) (Tao et al., 2009), Hepatitis C virus (ENC = 52.62), West Nile virus (ENC = 53.81) (Moratorio et al., 2013) and Dengue virus (DENV) (ENC = 49.70) (Ma et al., 2013). A possible explanation for a slight codon bias in these viruses is that it might be advantageous for efficient replication in host cells with potentially distinct codon preferences, as previously suggested (Jenkins and Holmes, 2003).

Since codon usage by its very nature is multivariate, it is necessary to analyze the data using different and complementary approaches. An ENC–$GC3_S$ plot (ENC plotted against $GC3_S$) can be used as a method that quantifies how far the codon usage of a gene departs from equal usage of synonymous codons (Wrigth, 1990). If $GC3_S$ is the only determinant factor shaping the codon usage pattern, the values of ENC would fall on a continuous curve, which represents random codon usage (Jiang et al., 2007). If G+C compositional constraint influences the codon usage, then the $GC3_S$ and ENC correlated spots would lie on or below the expected curve. When the ENC and$GC3_S$ values were calculated for ZEBOV ORFs and the ENC–$GC3_S$ plots constructed, all spots lie roughly below in relation to the expected curve, indicating that G+C compositional constraints might play a role in ZEBOV codon usage (see Fig. 1).

3.2. Trends of codon usage variation in ZEBOV

In order to gain insight into the trends in codon usage variation among different ZEBOV genomes, we performed a COA analysis. COA is a multivariate descriptive data analytic technique. This exploratory data analysis is used to identify systematic relations between variables when there are no (or rather incomplete) a priori expectations as to the nature of those relations. COA remarkably simplifies complex data and provides a detailed description of practically every bit of information in the data, yielding a simple, yet exhaustive analysis. It uses relationships between variables to order the objects of study according to their collective properties. The ultimate objective in this case is to relate the observed biological variation, explained by each axis. COA was performed on the RSCU values for each strain and examined the distribution of the strains along the plane determined by the first two principal axes of COA. The first axis generated by the analysis accounts for 72.03% of the total variation, while the second axis accounts for 12.25%. The results of these studies are shown in Fig. 2.

As it can be seen, the distribution of the ZEBOV isolates in the plane defined by the first two major axes showed that different isolates were located at different places, accordingly to their geographical origin of isolation (see Fig. 2). This may imply that the geographical diversity might play a role in shaping the molecular evolution and codon usage in ZEBOV, as well as the emergence of new ZEBOV strains. Moreover, these results may also imply that more than one genetic lineage can circulate in a single country (see Fig. 2). This is in agreement with recent results showing that ZEBOV from Guinea forms a separate clade in relation to the known ZEBOV strains from the Democratic Republic of Congo and Gabon (Baize et al., 2014).

These results are also in agreement with previous studies showing that ZEBOV has spread across the West African region in a wave-like manner rather than being long time persistent at each outbreak locality (Walsh et al., 2005).

3.3. Codon usage preferences in ZEBOV

In order to compare the codon usage preferences of ZEBOV with those of its human hosts, the RSCU values of the codons in ZEBOV ORFs were calculated and compared with those of human cells.

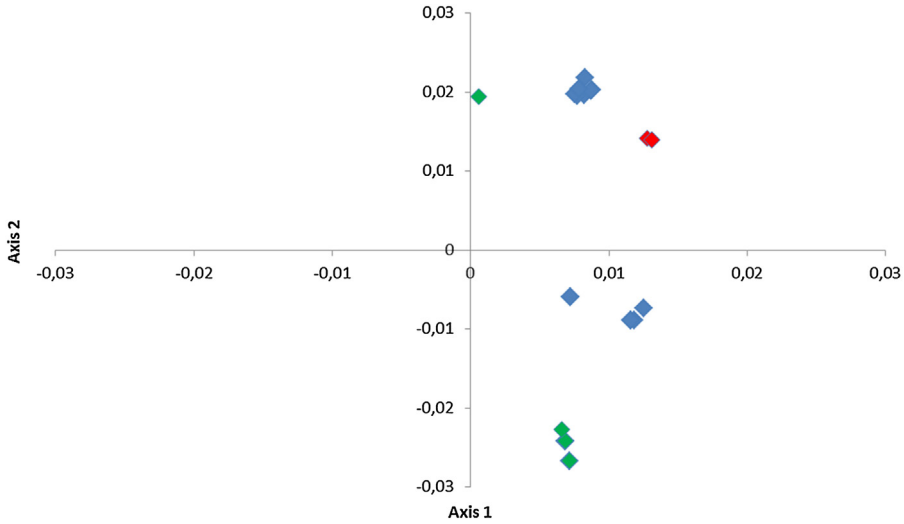

Fig. 2. Positions of the ZEBOV strains in the plot of the first two major axes by correspondence analysis (COA) of relative synonymous codon usage (RSCU) values. The first and second axes account for 72.03% and 12.25% of the total variation, respectively. The ZEBOV strains are divided according to their geographical origin of isolation. Strains isolated in Democratic Republic of Congo are indicated by blue diamonds, strains isolated in Gabon and the strains of the current 2014 outbreak isolated in Guinea are shown in green and red, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Codon usage in Zaire Ebolavirus, displayed as RSCU values.

AA	Cod	HC	EV	AA	Cod	HC	EV	AA	Cod	HC	EV	AA	Cod	HC	EV
Phe	UUU	0.92	0.95	Ser	UCU	1.14	1.17	Tyr	UAU	0.88	1.13	Cys	UGU	0.92	0.89
	UUC	1.08	1.05		UCC	1.32	0.95		UAC	1.12	0.87		UGC	1.08	1.11
Leu	**UUA**	**0.48**	**1.05**		**UCA**	**0.90**	**1.38**	TER	UAA	**	**		TER	UGA	**
	UUG	0.78	1.24		UCG	0.30	0.56		UAG	**	**		Trp	UGG	1.00
	CUU	0.78	0.97	Pro	CCU	1.16	0.89	His	CAU	0.84	1.08	Arg	CGU	0.48	0.61
	CUC	1.20	0.75		CCC	1.28	0.79		CAC	1.16	0.92		CGC	1.08	0.72
	CUA	**0.42**	**0.91**		**CCA**	**1.12**	**1.74**	Gln	**CAA**	**0.54**	**1.15**		CGA	0.66	0.75
	CUG	2.40	1.08		CCG	0.44	0.58		CAG	1.46	0.85		CGG	1.20	0.72
Ile	AUU	1.08	1.16	Thr	ACU	1.00	0.94	Asn	AAU	0.94	1.10	Ser	AGU	0.90	1.20
	AUC	1.40	1.08		ACC	1.44	1.09		AAC	1.06	0.90		AGC	1.44	0.73
	AUA	0.51	0.76		**ACA**	**1.12**	**1.52**	Lys	AAA	0.86	0.97	Arg	**AGA**	**1.26**	**2.01**
Met	AUG	1.00	1.00		ACG	0.44	0.44		AAG	1.14	1.03		AGG	1.26	1.19
Val	GUU	0.72	0.95	Ala	GCU	1.08	1.22	Asp	GAU	0.92	1.08	Gly	GGU	0.64	0.98
	GUC	0.96	0.92		GCC	1.60	0.86		GAC	1.08	0.92		GGC	1.36	0.79
	GUA	**0.48**	**0.87**		**GCA**	**0.92**	**1.46**	Glu	GAA	0.84	0.99		**GGA**	**1.00**	**1.47**
	GUG	1.84	1.26		GCG	0.44	0.46		GAG	1.16	1.01		GGG	1.00	0.76

RSCU, relative synonymous codon usage; AA, amino acid; Cod, codons; HC, human cells; EV, Ebola virus. Highly increased codons with respect to human cells ($\Delta \geq 0.30$) are shown in bold.
** Termination codons.

The results of these studies are shown in Table 1. Interestingly, the frequencies of codon usage in ZEBOV ORFs are significantly different in relation to human cells. Particularly, high biased frequencies were found for UUA and CUA (Leu), GUA (Val), UCA (Ser), CCA (Pro), ACA (Thr), CAA (Gln), GCA (Ala), AGA (Arg) and GGA (Gly) (see Table 1). As it can be seen, highly preferred codons are all A-ending, which strongly suggests that mutational bias is a main force shaping codon usage in ZEBOV. Previous studies have revealed that coincident portions of codon usage among virus and hosts may permit the corresponding amino acids to be translated efficiently, while the antagonistic portions of codon usage may enable viral proteins to be folded properly, even though the translation efficiency of the corresponding amino acids might decrease (Hu et al., 2014; Costafreda et al., 2014; Aragonès et al., 2010).

3.4. Codon usage adaptation in ZEBOV

In order to gain insight into the codon preferences of ZEBOV in relation of it human host, CAI values for all triplets were calculated, using human codon usage as a reference set. CAI index ranges from 0 to 1, being 1 if the frequency of codon usage by ZEBOV equals the frequency of usage of the reference set. A mean CAI of 0.705 were obtained for EBOV genes in relation to human codon usage reference set, while a mean CAI of 0.809 were obtained for a human genes dataset in relation to the same reference sample (see Table 2). In order to observe if the differences in CAI values between the two comparisons were statistically significant, we performed a Student's *t*-test and a Wilcoxon & Mann–Whitney test. The results of these tests revealed that the differences in CAI values are statistically significant ($t = 16.24$, p-value = 0 and $T = 625$, p-value = 0, respectively).

In order to discern if the statistically significant differences in CAI values arise from codon preferences (Puigbo et al., 2008a), the expected CAI (e-CAI) values were calculated for EBOV sequences in relation to human codon usage set. The e-CAI algorithm (Puigbo et al., 2008b) generated 500 random sequences with the same nucleotide and amino acid composition as the sequences of interest (in this case ZEBOV sequences), calculated the CAI values for all of them, and applied a Kolmogorov–Smirnov test for the e-CAI of these random sequences in order to show whether the generated sequences follow a normal distribution. The results of these studies revealed an e-CAI value of 0.741 ($p < 0.05$) and Kolmogorov–Smirnov test revealed a normal distribution of the generated sequences (Kolmogorov–Smirnov test of e-CAI value of 0.040, below of critical value of 0.061). Taking all these results together, our studies revealed that the CAI values for ZEBOV genes are different from the CAI values obtained from human ones. Again, these studies suggest that these differences are related to codon usage preferences.

When ZEBOV codons are sorted according to their position on the four major axes of COA, it is evident that the most extreme values are displayed by rarely used triplets, almost all of them containing the dinucleotide CpG (see Table 3). This fact reinforces the importance of CpG for the overall pattern of codon usage among ZEBOV strains.

It has been suggested that dinucleotide biases can affect codon bias (Tao et al., 2009). To study this possibility, the relative abundances of the 16 dinucleotides in ZEBOV ORFs were established. The results of these studies are shown in Table 4.

As can be seen, the occurrences of dinucleotides are not random and no dinucleotides are present at the expected frequencies. The relative abundance of CpG and GpC showed a strong deviation from the expected frequencies (i.e. 1.0) (mean ± S.D. = 0.45 ± 0.008 and 0.68 ± 0.044, respectively) and were markedly underrepresented. On the other hand, CpA and ApA are markedly over-used (mean ± S.D. = 1.35 ± 0.006 and 1.55 ± 0.009, respectively) (Table 4).

Previous studies have shown that CpG deficiency in pathogens is associated with the immunostimulatory properties of unmethylated CpGs, which are recognized by the host's innate immune

Table 2
Codon adaptation of EBOV genes in relation to Human codon usage, displayed as CAI values.

	CAI-Hs	%GC	%GC(1)	%GC(2)	%GC(3)
EBOV genes	0.705 ± 0.004	39.92 ± 1.25	45.51 ± 1.46	38.33 ± 0.69	35.46 ± 3.73
Human genes	0.809 ± 0.038	56.47 ± 7.72	58.37 ± 6.88	44.35 ± 7.71	66.69 ± 14.81

CAI, codon adaptation index; CAI-Hs, codon adaptation index in relation to Homo sapiens reference codon usage set. %GC, percentage of G+C genomic content, %GC(1) through (3), percentage of G+C genomic content in codon positions 1 through 3, respectively. In all cases, mean ± standard deviation values are shown.

Table 3
Position of codons in each of the four major axes of COA for EBOV proteins.

Axis 1			Axis 2			Axis 3			Axis 4		
Codon	Value	AA	Codon	Value	AA	Codon	Value	AA	Codon	Value	AA
GUU	−0.0959	Val	GCC	−0.04589	Ala	ACG	−0.05565	Thr	CGA	−0.04136	Arg
UCG	0.0848	Ser	CGU	0.06405	Arg	GCG	0.03744	Ala	CGG	0.01754	Arg

AA, amino acid.

Table 4
Relative abundance of dinucleotides in EBOV strains and summary of COA.

	UU	UC	UA	UG	CU	CC	CA	CG
Mean ± S.D.[a]	1.28 ± 0.0	1.00 ± 0.003	0.87 ± 0.008	1.05 ± 0.007	0.92 ± 0.008	0.78 ± 0.010	1.35 ± 0.006	0.45 ± 0.008
Axis 1[b]								
r	0.502	0.559	0.160	0.093	0.252	0.509	0.468	0.172
P	<0.05	<0.05	0.429	0.645	0.215	<0.05	<0.05	0.395
	AU	AC	AA	AG	GU	GC	GA	GG
Mean ± S.D.[a]	1.28 ± 0.007	1.03 ± 0.007	1.55 ± 0.009	1.06 ± 0.008	0.71 ± 0.010	0.68 ± 0.044	1.15 ± 0.006	0.77 ± 0.006
Axis 1[b]								
r	0.449	0.103	−0.315	0.088	0.368	0.449	0.112	0.348
P	<0.05	0.610	0.121	0.660	0.070	<0.05	0.575	0.087

[a] Mean values of EBOV strains relative dinucleotide ratios ± standard deviation.
[b] Correlation analysis between the first axis in COA and the sixteen dinucleotides frequencies in EBOV proteins is shown.

system as a pathogen signature (Shackelton et al., 2006). Recognition of unmethylated CpGs by Toll like receptor 9 (TLR9), leads to activation of several immune response pathways (Dorn and Kippenberger, 2008). The vertebrate immune system relies on unmethylated CpG recognition in DNA molecules as a signature of infection, and CpG under-representation has been observed in several RNA viruses; therefore, it is reasonable to suggest that a TLR9-like mechanism exists in the vertebrate immune system that recognizes CpGs when in an RNA context of the genomes of RNA viruses and triggers immune responses (Lobo et al., 2009). Over-representation of CpA has been observed in different organisms and is regarded as a consequence of the under-representation of CpG dinucleotides (Butt et al., 2014). Recent studies carried out in Human Rhinoviruses revealed that CpG suppression is counteracted by the over-representation of CpA. Furthermore, CpA-containing codons seem to function as "balancers" for the decrease of CpG-containing codons (Megremis et al., 2012).

Besides, among the 16 dinucleotides, six are correlated with the position of the sequences along the first axis in COA (*p* values < 0.05, Table 4). These results indicate that the composition of dinucleotides also determines the variation in synonymous codon usage among ZEBOV.

3.5. EBOV codon usage and cell tRNA isoacceptor abundance

The translation-selection hypothesis (Sharp et al., 2010) relies on the different concentration of isoacceptor tRNAs. As the translation process represents a key step in the viral infection cycle, it is important to explore the strategies employed by them to harness the translation machinery of the cell host. Since variation at the third codon position can change the wobble interaction between that base and the first one of the anticodon (Crick, 1966), we wanted to gain further insight into the adaptation of EBOV to the respective human cell tRNA pool context (see Table 5). Although the data presented here concerns to the number of each gene encoding a tRNA in the human genome, since the actual concentration of these tRNAs in the different human tissues that EBOV can infect is currently unknown, it gives a general picture on the relation of virus and host taking into consideration the relative abundance of tRNAs in the tRNA pool of a human cell. As it can be seen in Table 5, ZEBOV do not seem to use the most abundant tRNAs present in the human cells for their most preferred codons for Leu, Val, Ser, Pro, Thr, Ala, Gln, Arg and Gly, since no correlation is observed between the number of isoacceptor tRNAs and codon usage by the virus (see Table 5). This may suggest that the virus do not compete with the host genes for the most abundant pool of tRNAs during protein synthesis, at least for these amino acids.

Ribosomes translate messenger RNAs at a non-uniform speed which is shaped by several different factors, including codon usage, tRNA concentration and specificity of codon–anticodon interactions (Fedyunin et al., 2012). Different codons are read by tRNAs which largely vary in their abundance (Dong et al., 1996), and thus the ribosomes translate each codon with different speed. Fast-translated codons tend to dominate in highly expressed genes

Table 5
Frequency of tRNA genes in human cells for highly biased codons in EBOV.

AA	Cod	Anticodon isotypes(tRNA count by anticodon)	Total tRNA anticodon count
Leu	**UUA**, **CUA**	AGG(12), GAG(0), CAG(10), **TAG**(3), CAA(7), **TAA**(7)	39
Val	**GUA**	ACC(11), GAC(0), CAC(16), **TAC**(5)	32
Ser	**UCA**	ACA(11), GGA(0), CGA(4), **TGA**(5), ACT(0), GCT(8)	28
Pro	**CCA**	AGG(10), GGG(0), CGG(4), **TGG**(7)	21
Thr	**ACA**	AGT(10), GTT(0), CGT(6), **TGT**(6)	22
Ala	**GCA**	AGC(29), GGC(0), CGC(5), **TGC**(9)	43
Gln	**CAA**	CTG(20), **TTG**(11)	31
Arg	**AGA**	ACG(7), GCG(0), CCG(4), TCG(6), CTT(5), **TCT**(6)	28
Gly	**GGA**	ACC(0), GCC(15), CCC(7), **TCC**(9)	31

Highly biased codons in EBOV and their respective anticodons are shown in bold. AA, amino acid; Cod, codons.

(Jansen et al., 2003) and preferably encode buried residues most likely to reduce the frequency of translation errors at aggregation-prone sites (Lee et al., 2010). Clustering of slow-translating codons, i.e. that pair to low-abundant tRNAs, facilitate the hierarchical co-translational folding of single domains in multidomain proteins (Thanaraj and Argos, 1996). More studies will be needed in order to address these important issues in the case of ZEBOV.

4. Conclusions

The results of these studies revealed different codon preferences in ZEBOV genes in relation to codon usage of human genes (see Tables 1 and 2). The overall codon usage among ZEBOV strains is similar and slightly biased. G+C compositional constraint influences the codon usage of ZEBOV. Highly frequent codons are all A-ending, which strongly suggests that mutational bias is a main force shaping codon usage in this virus. Dinucleotide composition also plays a role in the overall pattern of ZEBOV codon usage. ZEBOV do not seem to use the most abundant tRNAs present in the human cells for their most preferred codons for Leu, Val, Ser, Pro, Thr, Ala, Gln, Arg and Gly, suggesting that the virus do not compete with the host cell for the most abundant pool of tRNAs during protein synthesis for these amino acids.

Acknowledgements

Authors acknowledge support by Agencia Nacional de Investigación e Innovación (ANII) through project PE_ALI_2009_1_1603, Fondo María Viñas and PEDECIBA, Uruguay. We also thank the support of Fondo Clemente Estable, 2007_722 to HM. We acknowledge anonymous reviewers for helpful suggestions to improve of the quality of this work.

References

Aragonès, L., Guix, S., Ribes, E., Bosch, A., Pintó, R.M., 2010. Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. PLoS Pathog. 6, e1000797.
Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., et al., 2014. Emergence of Zaire Ebola virus disease in Guinea—preliminary report. N. Engl. J. Med., http://dx.doi.org/10.1056/NEJMoa1404505.
Becquart, P., Mahlakoiv, T., Nkoghe, D., Leroy, E.M., 2014. Identification of continuous human B-cell epitopes in the VP35, VP40, nucleoprotein and glycoprotein of Ebola virus. PLoS ONE 9, e96360.
Burns, C.C., Shaw, J., Campagnoli, R., Jorba, J., Vincent, A., Quay, J., Kew, O., 2006. Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. J. Virol. 80, 3259–3272.
Butt, A.M., Nasrullah, I., Tong, Y., 2014. Genome-wide analysis of codon usage and influencing factors in Chikungunya viruses. PLoS ONE 9, e90905.
Centers for Disease Control, 2001. Ebola outbreak in Uganda. Morb. Mortal. Wkly. Rep. 50, 73–77.
Chan, P.P., Lowe, T.M., 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res. 37, D93–D97.
Costafreda, M.I., Pérez-Rodriguez, F.J., D'Andrea, L., Guix, S., Ribes, E., Bosch, A., Pintó, R.M., 2014. Hepatitis A virus adaptation to cellular shutoff is driven by dynamic adjustments of codon usage and results in the selection of populations with altered capsids. J. Virol. 88, 5029–5041.
Crick, F.H., 1966. Codon–anticodon pairing: the wobble hypothesis. J. Mol. Biol. 19, 548–555.
Dixon, M.G., Schafer, I.J., 2014. Ebola viral disease outbreak—West Africa, 2014. Morb. Mortal. Wkly. Rep. 63 (June (25)), 548–551.
Dong, H., Nilsson, L., Kurland, C.G., 1996. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. J. Mol. Biol. 260, 649–663.
Dorn, A., Kippenberger, S., 2008. Clinical application of CpG-, non-CpG, and antisense oligodeoxynucleotides as immunomodulators. Curr. Opin. Mol. Ther. 10, 10–20.
Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform. 5, 113.
Feldmann, H., Klenk, H.D., Sanchez, A., 1993. Molecular biology and evolution of filoviruses. Arch. Virol. Suppl. 7, 81–100.
Fedyunin, I., Lehnhardt, L., Böhmer, N., Kaufmann, P., Zhang, G., Ignatova, Z., 2012. tRNA concentration fine tunes protein solubility. FEBS Lett. 586, 3336–3340.
Frieden, T.R., Damon, I., Bell, B.P., Kenyon, T., Nichol, S., 2014. Ebola 2014—new challenges, new global response and responsibility. N. Engl. J. Med., http://dx.doi.org/10.1056/NEJMp1409903.
Geisbert, T.W., Hensley, L., Larsen, T., Young, H.A., Reed, D.S., et al., 2003. Pathogenesis of Ebola hemorrhagic fever in cynomolgus macaques. Am. J. Pathol. 163, 2347–2370.
Gingold, H., Pilpel, Y., 2011. Determinants of translation efficiency and accuracy. Mol. Syst. Biol. 7, 481.
Gire, S.K., Goba, A., Andersen, K.G., Sealfon, R.S., Park, D.J., et al., 2014. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science 345, 1369–1372.
Greenacre, M., 1984. Theory and Applications of Correspondence Analysis. Academic Press, London.
Hoenen, T., Groseth, A., Falzarano, D., Feldmann, H., 2006. Ebola virus: unravelling pathogenesis to combat a deadly disease. Trends Mol. Med. 12, 206–215.
Hu, C., Chen, J., Ye, L., Chen, R., Zhang, L., Xue, X., 2014. Codon usage bias in human cytomegalovirus and its biological implication. Gene 545, 5–14.
Jansen, R., Bussemaker, H.J., Gerstein, M., 2003. Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. Nucleic Acids Res. 31, 2242–2251.
Jenkins, G.M., Holmes, E.C., 2003. The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res. 92, 1–7.
Jiang, P., Sun, X., Lu, Z., 2007. Analysis of synonymous codon usage in *Aeropyrum pernix* K1 and other *Crenarchaeota* microorganisms. J. Genet. Genomics 34, 275–284.
Khan, A.S., Tshioko, F.K., Heymann, D.L., Le Guenno, B., Nabeth, P., et al., 1999. The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo 1995. J. Infect. Dis. 179, S76–S86.
Lee, Y., Zhou, T., Tartaglia, G.G., Vendruscolo, M., Wilke, C.O., 2010. Translationally optimal codons associate with aggregation-prone sites in proteins. Proteomics 10, 4163–4171.
Le Guenno, B., Formenty, P., Boesch, C., 1999. Ebola virus outbreaks in the Ivory Coast and Liberia, 1994–1995. Curr. Top. Microbiol. Immunol. 235, 77–84.
Lobo, F.P., Mota, B.E., Pena, S.D., Azevedo, V., Macedo, A.M., et al., 2009. Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. PLoS ONE 20, e6282.
Ma, J.J., Zhao, F., Zhang, J., Zhou, J.H., Ma, L.N., et al., 2013. Analysis of synonymous codon usage in Dengue viruses. J. Anim. Vet. Adv. 12, 88–98.
Megremis, S., Demetriou, P., Makrinioti, H., Manoussaki, A.E., Papadopoulos, N.G., 2012. The genomic signature of Human Rhinoviruses A, B and C. PLoS ONE 7, e44557.
Moratorio, G., Iriarte, A., Moreno, P., Musto, H., Cristina, J., 2013. A detailed comparative analysis on the overall codon usage patterns in West Nile virus. Infect. Genet. Evol. 14, 396–400.
Mueller, S., Papamichail, D., Coleman, J.R., Skiena, S., Wimmer, E., 2006. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. J. Virol. 80, 9687–9696.
Novembre, J.A., 2002. Accounting for background nucleotide composition when measuring codon usage bias. Mol. Biol. Evol. 19, 1390–1394.
Puigbo, P., Bravo, I.G., Garcia-Vallve, S., 2008a. CAIcal: a combined set of tools to assess codon usage adaptation. Biol. Direct 3, e38.
Puigbo, P., Bravo, I.G., Garcia-Vallve, S., 2008b. E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). BMC Bioinform. 9, e65.
Rasmussen, A.L., Okumura, A., Ferris, M.T., Green, R., Feldmann, F., et al., 2014. Host genetic diversity enables Ebola hemorrhagic fever pathogenesis and resistance. Science, http://dx.doi.org/10.1126/science.1259595.
Rollin, P.E., Williams, R.J., Bressler, D.S., Pearson, S., Cottingham, M., et al., 1999. Ebola (subtype Reston) virus among quarantined nonhuman primates recently imported from the Philippines to the United States. J. Infect. Dis. 179, S108–S114.
Stroher, U., West, E., Bugany, H., Klenk, H.D., Schnittler, H.J., et al., 2001. Infection and activation of monocytes by Marburg and Ebola viruses. J. Virol. 75, 11025–11033.
Shackelton, L.A., Parrish, C.R., Holmes, E.C., 2006. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J. Mol. Evol. 62, 551–563.
Sharp, P.M., Emery, L.R., Zeng, K., 2010. Forces that influence the evolution of codon bias. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 365, 1203–1212.
Sharp, P.M., Li, W.H., 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15, 1281–1295.
Sharp, P.M., Li, W.H., 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. 24, 28–38.
Sanchez, A., Geisbert, T.W., Feldmann, H., 2007. Filoviridae: Marburg and Ebola viruses. In: Knipe, D.M., Howley, P.M. (Eds.), Fields Virology. Lippincott Williams and Williams, Philadelphia, pp. 1409–1448.
Stoletzki, N., Eyre-Walker, A., 2007. Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. Mol. Biol. Evol. 24, 374–381.
Thanaraj, T.A., Argos, P., 1996. Ribosome-mediated translational pause and protein domain organization. Protein Sci. 5, 1594–1612.
Tao, P., Dai, L., Luo, M., Tang, F., Tien, P., Pan, Z., 2009. Analysis of synonymous codon usage in classical swine fever virus. Virus Genes 38, 104–112.
Towner, J.S., Sealy, T.K., Khristova, M.L., Albariño, C.G., Conlan, S., et al., 2008. Newly discovered Ebola virus associated with hemorrhagic fever outbreak in Uganda. PLoS Pathog. 4, e1000212.
Walsh, P.D., Biek, R., Real, L.A., 2005. Wave-like spread of Ebola Zaire. PLoS Biol. 3, e371.

Wang, M., Zhang, J., Zhou, J.H., Chen, H.T., Ma, L.N., et al., 2011. Analysis of codon usage in bovine viral diarrhea virus. Arch. Virol. 156, 153–160.
Wauquier, N., Becquart, P., Padilla, C., Baize, S., Leroy, E.M., 2010. Human fatal Zaire Ebola Virus infection is associated with an aberrant innate immunity and with massive lymphocyte apoptosis. PLoS Negl. Trop. Dis. 4, e837.
Wessa, P., 2012. Free Statistics Software, Office for Research Development and Education, Version 1.1.23-r7. http://www.wessa.net
Wong, E.H., Smith, D.K., Rabadan, R., Peiris, M., Poon, L.L., 2010. Codon usage bias and the evolution of influenza A viruses. Codon usage biases of influenza Virus. BMC Evol. Biol. 19, 253.
World Health Organization, 2004. Ebola haemorrhagic fever in south Sudan—update. Wkly. Epidemiol. Rec. 79, 253.
Wrigth, F., 1990. The “effective number of codons” used in a gene. Gene 87, 23–29.