

SHORT REPORT

Open Access

# Comprehensive analysis of the overall codon usage patterns in equine infectious anemia virus

Xin Yin<sup>1,2</sup>, Yuezhi Lin<sup>1</sup>, Weigang Cai<sup>1</sup>, Ping Wei<sup>2\*</sup> and Xiaojun Wang<sup>1\*</sup>

## Abstract

**Background:** Equine infectious anemia virus (EIAV) is an important animal model for understanding the relationship between viral persistence and the host immune response during lentiviral infections. Comparison and analysis of the codon usage model between EIAV and its hosts is important for the comprehension of viral evolution. In our study, the codon usage pattern of EIAV was analyzed from the available 29 full-length EIAV genomes through multivariate statistical methods.

**Finding:** Effective number of codons (ENC) suggests that the codon usage among EIAV strains is slightly biased. The ENC-plot analysis demonstrates that mutation pressure plays a substantial role in the codon usage pattern of EIAV, whereas other factors such as geographic distribution and host translation selection also take part in the process of EIAV evolution. Comparative analysis of codon adaptation index (CAI) values among EIAV and its hosts suggests that EIAV utilize the translational resources of horse more efficiently than that of donkey.

**Conclusion:** The codon usage bias in EIAV is slight and mutation pressure is the main factor that affects codon usage variation in EIAV. These results suggest that EIAV genomic biases are the result of the co-evolution of genome composition and the ability to evade the host's immune response.

**Keywords:** Equine infectious anemia virus (EIAV), Codon usage bias, Evolution

## Findings

### Introduction

Equine infectious anemia virus (EIAV) is an important non-primate enveloped virus, of the retrovirus family, lentivirus genus, along with the human immunodeficiency virus (HIV), simian immunodeficiency virus (SIV) [1]. Among the lentiviruses, EIAV is the least complex lentivirus including only 6 genes. In addition to the *gag*, *pol* and *env* genes coding for the structural and enzymatic proteins coded by *gag*, *pol* and *env*, EIAV also contains three accessory genes: *tat*, *rev* and *S2* [2]. The host range of EIAV is reported to include all members of the *Equidae*, while susceptible to infection, donkeys do not develop clinical EIA and lower amounts of plasma associated virus [3].

It is well known that the redundancy of the genetic code allows for multiple codons to encode for a single amino

acid, resulting in codon usage biases in genes [4]. The non-random usage of synonymous codons is crucial for the efficient protein translation and correct folding. Indeed, mutation pressure and natural selection are thought to be two major forces that drive the codon usage bias away from an equal usage among genes in different organisms [5]. Understanding the extent and causes of biases in codon usage is important for the comprehension of the pathogen evolution and the relationship between pathogens and the immune response [6].

Recent efforts to understand codon usage biases in viruses have primarily focused on the hepatitis A virus [7,8], West Nile virus [9], foot-and-mouth disease virus [10], influenza virus [11], and HIV [12-14]. To date, although the remarkable adenine (A)-richness of the EIAV genome was already discovered several decades ago [15], few codon usage analyses have been performed on EIAV genome. To gain insight into the characteristics of the viral genome, the synonymous codon usage pattern and the correlation between the codon usage pattern of EIAV and its hosts were investigated in our study.

\* Correspondence: weiping@yahoo.com.cn; xjw@hvri.ac.cn

<sup>2</sup>College of Veterinary Medicine, Northeast Agricultural University, Harbin 150030, PR China

<sup>1</sup>Division of Livestock Infectious Diseases, State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin 150001, PR China

**Table 1 The overall nucleotide contents and nucleotide contents at the synonymous third position of sense codons in EIAV genome**

No. of genomes analyzed	T (%) ± std	C (%) ± std	A (%) ± std	G (%) ± std	T 3(%) ± std	C3 (%) ± std	A3 (%) ± std	G3 (%) ± std
29	25.39 ± 0.32	15.57 ± 0.33	37.09 ± 0.32	21.96 ± 0.24	29.62 ± 0.78	11.52 ± 0.58	38.01 ± 0.83	20.85 ± 0.63

## Methods

### Sequence data

The complete genome sequences of 29 EIAV strains were obtained from the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/Genbank/>). The detailed information about the viruses is listed in Additional file 1: Table S1.

### Codon usage analysis

Each nucleotide content and each nucleotide content at the third site of the codon in the EIAV coding sequence were calculated using MEGA4 software. The dinucleotides of the EIAV genome were analyzed by DAMBE software. The relative synonymous codon usage (RSCU) values for EIAV were calculated as previously described [16]. The effective

**Table 2 Codon usage in EIAV genomes and its hosts**

AA	Codon	RSCU			AA	Codon	RSCU		
		EIAV	Horse	Donkey			EIAV	Horse	Donkey
Phe	UUU	1.48	0.82	0.89	Cys	UGU	1.50	0.90	0.62
	UUC	0.52	1.18	1.11		UGC	0.50	1.10	1.38
Asn	AAU	1.50	0.84	0.66	Arg	CGU	0.17	0.36	0.48
	AAC	0.50	1.16	1.34		CGC	0.18	1.14	0.78
Lys	AAA	1.54	0.80	0.78	CGA	CGA	0.20	0.60	0.72
	AAG	0.46	1.20	1.22		CGG	0.31	1.08	0.60
Leu	UUA	2.192.06	0.30	0.24	AGA	AGA	3.09	1.30	1.50
	UUG	1.29	0.70	0.84		AGG	2.05	1.34	1.86
	CUU	0.58	0.74	0.78	Ser	UCU	1.42	1.08	1.08
	CUC	0.46	1.32	1.62		UCC	0.73	1.44	1.44
	CUA	0.85	0.36	0.18		UCA	1.41	0.78	0.84
CUG	0.62	2.58	2.34	UCG	0.12	0.48	0.24		
Pro	CCU	2.04	1.20	0.84	AGU	AGU	1.71	0.84	1.08
	CCC	0.40	1.28	1.60		AGC	0.61	1.50	1.26
	CCA	1.44	0.96	1.04	Asp	GAU	1.35	0.84	0.87
CCG	0.12	0.44	0.52	GAC		0.65	1.16	1.13	
Thr	ACU	1.62	0.92	0.84	Glu	GAA	1.25	0.76	0.84
	ACC	0.42	1.60	1.76		GAG	0.75	1.24	1.16
	ACA	1.79	0.96	0.80	His	CAU	1.45	0.80	0.88
	ACG	0.17	0.52	0.60		CAC	0.55	1.20	1.12
Val	GUU	0.74	0.60	0.64	Gln	CAA	1.28	0.52	0.84
	GUC	0.37	1.08	1.40		CAG	0.72	1.48	1.16
	GUA	1.96	0.36	0.28	Gly	GGU	0.64	0.64	0.88
	GUG	0.93	1.96	1.64		GGC	0.38	1.44	1.44
Ala	GCU	1.72	1.08	1.20	GGA	GGA	2.05	0.94	0.82
	GCC	0.44	1.72	1.76		GGG	0.92	0.98	0.86
	GCA	1.60	0.76	0.76	Ile	AUU	1.04	0.93	0.57
	GCG	0.24	0.44	0.28		AUC	0.45	1.65	1.95
Tyr	UAU	1.45	0.76	0.64	AUA	AUA	1.51	0.52	0.48
	UAC	0.55	1.24	1.36					

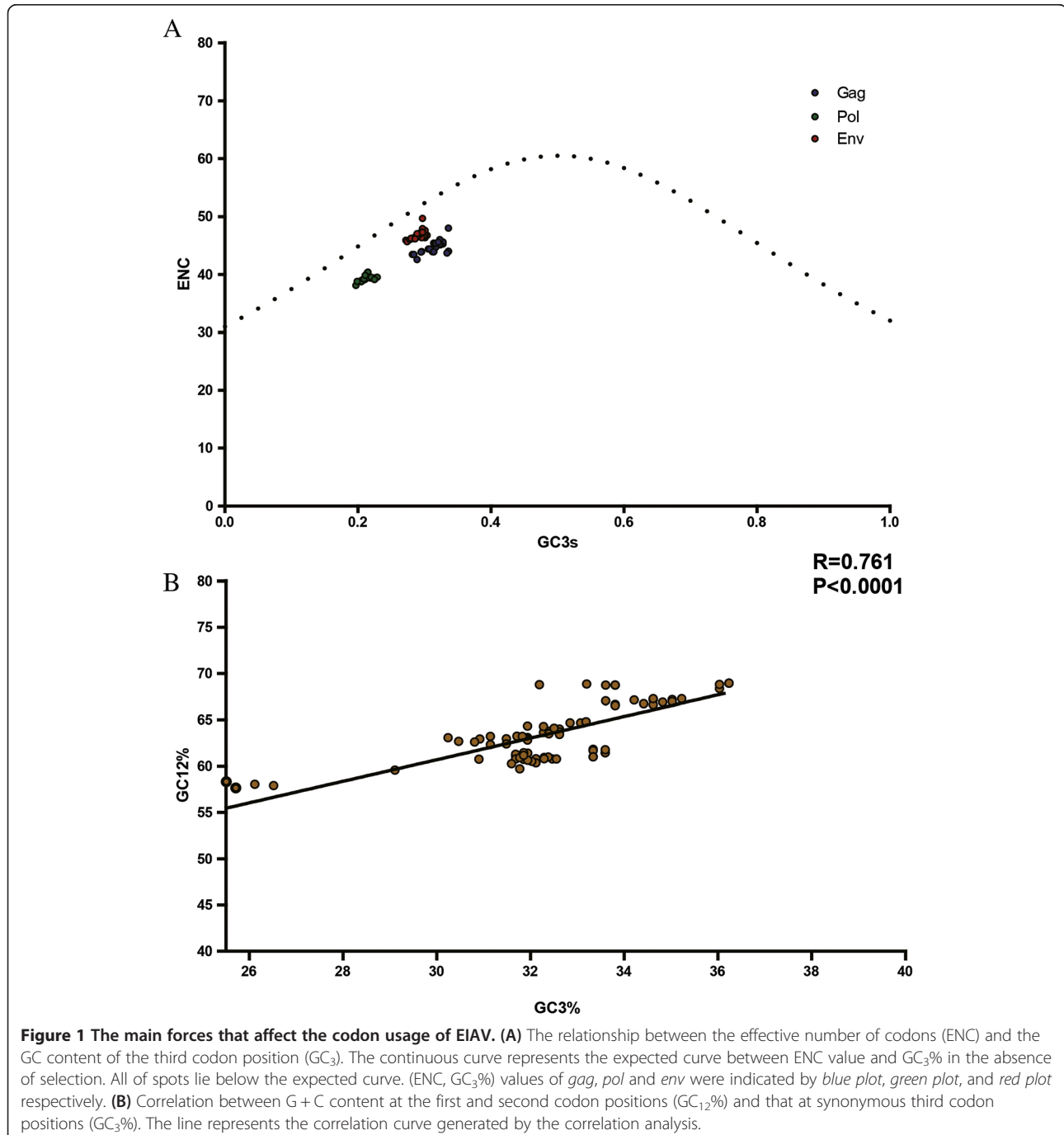
number of codons (ENC), was used to quantify deviations from the expected random codon usage of EIAV ORFs [17]. The ENC values range from 20 to 61, and a low ENC value indicates a strong codon usage bias.

The codon adaptation index (CAI) was used to estimate the adaptation of EIAV to host codons. When the CAI value is much closer to 1, the gene expression level is much higher. The CAI was calculated to compare a given codon usage to a predefined

reference set, using the CAIcal approach (available at: <http://genomes.urv.es/CAIcal>). The synonymous codon usage data for the viral hosts were obtained from the codon usage database (<http://www.kazusa.or.jp/codon/>) [18].

#### Principal component analysis

Principal component analyses (PCA) were performed to analyze the major trend in the codon usage model among

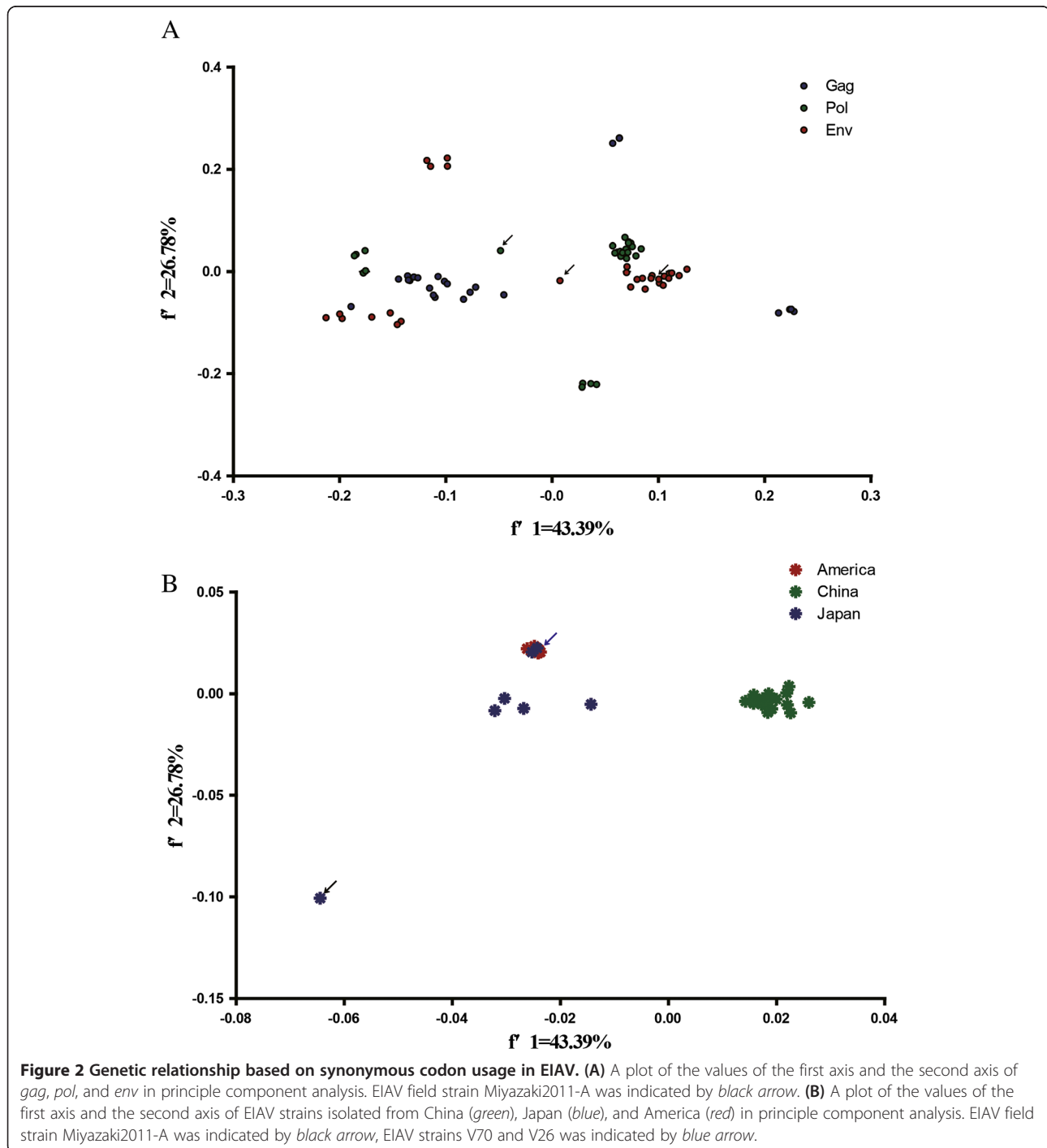


the different EIAV strains. Each ORF is represented as a 59-dimensional vector and each dimension corresponds to the RSCU value of one sense codon, excluding the codons of AUG, UGG and terminal codons. The major trend within a dataset can be determined using measure of relative inertia and genes ordered according to their position along the axis of major inertia [19].

## Results and discussion

### Synonymous codon usage in EIAV

The overall base composition of different EIAV strains was nonrandom, and the U% and A% were higher than the C% and G% (Table 1). The EIAV genome which was rich in A may possess viral tactics for escaping from the antiviral activity of apolipoprotein B mRNA-editing enzyme-catalytic



**Figure 2 Genetic relationship based on synonymous codon usage in EIAV. (A)** A plot of the values of the first axis and the second axis of *gag*, *pol*, and *env* in principle component analysis. EIAV field strain Miyazaki2011-A was indicated by *black arrow*. **(B)** A plot of the values of the first axis and the second axis of EIAV strains isolated from China (*green*), Japan (*blue*), and America (*red*) in principle component analysis. EIAV field strain Miyazaki2011-A was indicated by *black arrow*, EIAV strains V70 and V26 was indicated by *blue arrow*.

polypeptide 3 (APOBEC3) [20,21]. To investigate whether these 29 EIAV strains display similar codon usage biases, the ENC values were calculated. The obtained ENC values varied from 38.10 to 49.71 with a mean of  $43.61 \pm 3.28$ . One possible explanation for the weak synonymous codon bias of EIAV was that the weak codon bias is essential to increasing the translational accuracy and efficiency. To further analyze the extent of codon usage bias in EIAV, the overall RSCU values for the 59 sense codons were calculated. Almost all extremely highly preferred codons among the strains were A-ended or U-ended codons (Table 2). In addition, AA was the most common dinucleotide in EIAV, while CG dinucleotides were significantly suppressed in the genome of EIAV. The result revealed that CpG suppression may be a factor that shapes the synonymous codon usage of EIAV genome.

#### The effect of mutation pressure on the codon usage of EIAV

The ENC-plot analysis (ENC value plotted against the  $GC_{3s}$  content) was performed to further investigate the patterns of synonymous codon usage. We found that all of the spots lie slightly below the expected curve, indicating that mutational pressure was the main factor for shaping the codon usage bias of EIAV (Figure 1A). To further identify the role of mutation pressure from the virus itself or by natural selection pressure in shaping the codon usage pattern of EIAV, a correlation analysis was used to analyze the relationships among the G + C content at the first and second codon positions ( $GC_{12}\%$ ) and that at the synonymous third codon positions ( $GC_3\%$ ). A highly significant correlation was observed ( $r = 0.761$ ,  $P < 0.001$ ), indicating that the mutation pressure dominated over the natural selection pressure in shaping the coding sequence's composition (Figure 1B).

#### Genetic relationship based on synonymous codon usage in EIAV

The first principal component ( $f_1$ ) which can account for 43.39% of the total variation has a substantial impact on the total variation in the codon usage pattern. In addition, a plot of the  $f_1$  and  $f_2$  of the *gag*, *pol*, and *env* ORFs in EIAV was drawn (Figure 2A). The plots for the different structural proteins were generally separated from each other. This phenomenon implied that the functions of the viral protein were likely related to the codon usage pattern.

It has been reported that a strong pattern of geographic clustering is observed for EIAV, with a significant correlation between phylogroups of isolates and major geographic regions [22]. Based on the potential for the geographical factors in influencing EIAV evolution, a plot of  $f_1$  and  $f_2$  was performed according to the geographic distribution. The plots for EIAV isolated from China, Japan,

and America were generally divided into three groups, implying that the EIAV isolated from the three countries evolved independently after diverging from a common ancestor (Figure 2B). In addition, we cannot ignore that the plots for EIAV strains V70 and V26 were clustered together with the strains isolated from America. The origin of these strains still remains controversial [23,24]. Our data demonstrated that these EIAV strains have an American ancestry. Notably, the EIAV Miyazaki2011-A plot was far from the plots of the other strains. Recent reports showed that this EIAV strain was unlikely derived as a result of genomic recombination events and constituted a separate monophyletic group [24]. It is interesting to identify the potential origin of this novel EIAV isolate.

#### Comparative analysis of the codon usage between EIAV and host cells

The synonymous codon usage pattern of EIAV tended to differ from that of horse and donkey (Table 2 and Additional file 2: Figure S1). To further investigate whether the frequency of codon usage between EIAV and its hosts might have a close relationship with the viral proteins' expression levels, the CAI were calculated using the horse and donkey codon usage as reference sets [25]. A mean CAI of  $0.655 \pm 0.020$  was obtained for the EIAV ORFs in relation to horse codon usage reference set. A mean CAI of  $0.593 \pm 0.021$  was obtained for the EIAV ORFs in relation to the donkey codon usage reference set. There was a trend for a lower CAI for EIAV in relation to donkey, with the consequent lower efficiency of protein synthesis in donkey. This phenomenon reflected that the interplay of codon usage between EIAV and its hosts may influence viral fitness, survival and evolution.

In conclusion, our comprehensive analysis of the codon usage patterns in EIAV has provided a basic understanding about some of the evolutionary information of EIAV. However, there were some limitations to this study. The sample size was relatively small and may not be fully representative of EIAV. More studies should be carried out to confirm the conjecture.

#### Additional files

**Additional file 1: Table S1.** List of the information about the 29 EIAV genomes used in this study.

**Additional file 2: Figure S1.** The ENC-plot for mean codon usage in horse and donkey. The continuous curve represents the expected curve between ENC value and  $GC_3\%$  in the absence of selection. (ENC,  $GC_3\%$ ) values of horse and donkey were indicated by red plot and blue plot respectively.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

XY performed the experiments and wrote the first draft of the paper, in collaboration with XJW and PW. YZL and WGC analyzed the data and drafted the manuscript. All authors read and approved the final manuscript.

#### Acknowledgment

This work was supported by the Natural Science Foundation of China (Grant No. 31072113), the National Science Foundation for Outstanding Young Scholars of China (Grant No. 31222054), State Key Laboratory of Veterinary Biotechnology (Grant No. SKLVBP201205), and Central Public-interest Scientific Institution Basal Research Fund (Grant No. 2013ZL034).

Received: 18 October 2013 Accepted: 11 December 2013

Published: 20 December 2013

#### Reference

1. Leroux C, Cadore JL, Montelaro RC: Equine Infectious Anemia Virus (EIAV): what has HIV's country cousin got to tell us? *Vet Res* 2004, **35**:485–512.
2. Craigo JK, Montelaro RC: EIAV envelope diversity: shaping viral persistence and encumbering vaccine efficacy. *Curr HIV Res* 2010, **8**:81–86.
3. Cook SJ, Cook RF, Montelaro RC, Issel CJ: Differential responses of *Equus caballus* and *Equus asinus* to infection with two pathogenic strains of equine infectious anemia virus. *Vet Microbiol* 2001, **79**:93–109.
4. Hershberg R, Petrov DA: Selection on codon bias. *Annu Rev Genet* 2008, **42**:287–299.
5. Karlin S, Mrazek J: What drives codon choices in human genes? *J Mol Biol* 1996, **262**:459–472.
6. Shackelton LA, Holmes EC: The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends Microbiol* 2004, **12**:458–465.
7. Ma XX, Feng YP, Chen L, Zhao YQ, Liu JL, Guo JZ, Guo PH, Yang JT, Lu JX, Chen SE, Ma ZR: Mapping codon usage in sequence regions flanking cleavage positions in the hepatitis A virus polyprotein. *Genet Mol Res* 2013, **12**:2306–2319.
8. Zhang Y, Liu Y, Liu W, Zhou J, Chen H, Wang Y, Ma L, Ding Y, Zhang J: Analysis of synonymous codon usage in hepatitis A virus. *Virology* 2011, **8**:174.
9. Moratorio G, Iriarte A, Moreno P, Musto H, Cristina J: A detailed comparative analysis on the overall codon usage patterns in West Nile virus. *Infect Genet Evol* 2013, **14**:396–400.
10. Zhou JH, You YN, Chen HT, Zhang J, Ma LN, Ding YZ, Pejsak Z, Liu YS: The effects of the synonymous codon usage and tRNA abundance on protein folding of the 3C protease of foot-and-mouth disease virus. *Infect Genet Evol* 2013, **16**:270–274.
11. Goni N, Iriarte A, Comas V, Sonora M, Moreno P, Moratorio G, Musto H, Cristina J: Pandemic influenza A virus codon usage revisited: biases, adaptation and implications for vaccine strain development. *Virology* 2012, **9**:263.
12. van der Kuyl AC, Berkhout B: The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus. *Retrovirology* 2012, **9**:92.
13. Pandit A, Sinha S: Differential trends in the codon usage patterns in HIV-1 genes. *PLoS One* 2011, **6**:e28889.
14. Kypr J, Mrazek J: Unusual codon usage of HIV. *Nature* 1987, **327**:20.
15. van Hemert FJ, Berkhout B: The tendency of lentiviral open reading frames to become A-rich: constraints imposed by viral genome organization and cellular tRNA availability. *J Mol Evol* 1995, **41**:132–140.
16. Sharp PM, Li WH: An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol* 1986, **24**:28–38.
17. Comeron JM, Aguade M: An evaluation of measures of synonymous codon usage bias. *J Mol Evol* 1998, **47**:268–274.
18. Puigbo P, Bravo IG, Garcia-Vallve S: E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics* 2008, **9**:65.
19. Tao P, Dai L, Luo M, Tang F, Tien P, Pan Z: Analysis of synonymous codon usage in classical swine fever virus. *Virus Genes* 2009, **38**:104–112.
20. Zielonka J, Bravo IG, Marino D, Conrad E, Perkovic M, Battenberg M, Cichutek K, Munk C: Restriction of equine infectious anemia virus by equine APOBEC3 cytidine deaminases. *J Virol* 2009, **83**:7547–7559.
21. Bogerd HP, Tallmadge RL, Oaks JL, Carpenter S, Cullen BR: Equine infectious anemia virus resists the antiretroviral activity of equine APOBEC3 proteins through a packaging-independent mechanism. *J Virol* 2008, **82**:11889–11901.
22. Capomaccio S, Cappelli K, Cook RF, Nardi F, Gifford R, Marenzoni ML, Passamonti F: Geographic structuring of global EIAV isolates: a single origin for New World strains? *Virus Res* 2012, **163**:656–659.
23. Zheng YH, Sentsui H, Kono Y, Ikuta K: Mutations occurring during serial passage of Japanese equine infectious anemia virus in primary horse macrophages. *Virus Res* 2000, **68**:93–98.
24. Dong JB, Zhu W, Cook FR, Goto Y, Horii Y, Haga T: Identification of a novel equine infectious anemia virus field strain isolated from feral horses in southern Japan. *J Gen Virol* 2013, **94**:360–365.
25. Puigbo P, Bravo IG, Garcia-Vallve S: CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct* 2008, **3**:38.

doi:10.1186/1743-422X-10-356

Cite this article as: Yin et al.: Comprehensive analysis of the overall codon usage patterns in equine infectious anemia virus. *Virology Journal* 2013 **10**:356.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

