

## Phylogenetic Analysis of Coxsackie B Viruses Reveals Genomic Plasticity and Adaptation as Studied by Codon Usage Patterns

Ricardo Recarey<sup>1</sup>, Gonzalo Moratorio<sup>1,2</sup>, Rodney Colina<sup>3</sup>, Mónica Cappetta<sup>4,5</sup>, Rosario Uriarte<sup>4</sup>, Héctor Musto<sup>6</sup> and Juan Cristina<sup>1\*</sup>

<sup>1</sup>Laboratorio de Virología Molecular, Centro de Investigaciones Nucleares, Facultad de Ciencias, Universidad de la Republica, Uruguay

<sup>2</sup>Unidad de Biofísica de Proteínas, Instituto Pasteur de Montevideo, Uruguay

<sup>3</sup>Laboratorio de Virología Molecular, Regional Norte, Sede Salto, Universidad de la Republica, Uruguay

<sup>4</sup>Laboratorio de Biología Molecular, Asociación Española Primera de Socorros Mutuos, Uruguay

<sup>5</sup>Departamento de Genética, Facultad de Medicina, Universidad de la Republica, Uruguay

<sup>6</sup>Laboratorio de Organización y Evolución del Genoma, Instituto de Biología, Facultad de Ciencias, Universidad de la Republica, Uruguay

### Abstract

Coxsackie B viruses (CVB) are associated with serious illnesses in humans. In this study, the patterns of synonymous codon usage in CVB have been studied through multivariate statistical methods. Effective number of codons (ENC) indicates that the overall extent of codon usage bias in CVB is not significant. The relative dinucleotide abundances suggest that codon usage bias in CVB genomes is influenced by underlying biases of dinucleotide frequencies. The distribution of CVB ORFs along the plane defined by the first two axes of correspondence analysis (COA) showed that different genotypes, as well as strains known to infect different cell types, are located at different places in the plane suggesting that CVB codon usage is reflecting an evolutionary process. The results of these studies suggest that CVB genomic biases are the result of co-evolution of translation adaptation to different cell environments and probably the need to escape anti-viral cell defenses.

**Keywords:** Coxsackie B virus; Evolution; Codon usage; Adaptation

### Introduction

The genus *Enterovirus* of the family *Picornaviridae* are among the most common human viral pathogens, with more than 50 serotypes of enteroviruses that cause illness in humans [1]. Enteroviruses are the most common cause of aseptic meningitis, which is the most frequent central nervous system infection worldwide [2]. There is an increasing concern of the impact of this disease in South America [3,4]. Coxsackie viruses belong to this genus of the family and are divided into serogroups A and B, including six types in the B serogroup [5]. While Coxsackie A viruses are most commonly associated with skin exanthema, the Coxsackie B viruses (CVB) are associated with serious illnesses such as myocarditis, pericarditis, meningitis, diabetes, among other diseases [6,7]. All six types of CVB bind and enter cells through a common receptor protein, the coxsackie and adenovirus receptor (CAR) [8,9].

Due to the degeneracy of the genetic code, most amino acids are coded by more than one codon (synonymous codon usage), and those that code for the same amino acid have been observed to be used unequally in most species. Two major models have been proposed to explain codon usage, the translation related (or selective) model and the mutational (or neutral) model. In the translation related model, there is a co-adaptation of synonymous codon usage and tRNA abundance to optimize translational efficiency, and a correlation between codon usage and gene expression is found at levels of speed and accuracy [10]. In the mutational model, genomic compositional constraints influence the probability of mutational fixation. This has been found in many species [11,12]. However, these two models are not mutually exclusive [10].

Understanding the extent and causes of biases in codon usage is essential to the comprehension of viral evolution, particularly the interplay between viruses and the host cell, as well as between viruses and the immune response [13].

In order to gain insight into the processes governing the evolution and host adaptation of CVB, we have performed a detailed analysis

of codon usage biases of CVB strains, including strains isolated from Uruguayan children with diagnosis of aseptic meningitis, as well as strains isolated elsewhere from different patients and disease conditions. The results of these studies provide clues into the variation of codon usage pattern among the CVB genomes and the genomic plasticity of CVB in relation to adaptation to different cell hosts.

### Materials and Methods

#### Patients

Cerebrospinal fluid (CSF) samples from Uruguayan children with diagnosis of aseptic meningitis were obtained from Hospital Asociación Española Primera de Socorros Mutuos, Montevideo, Uruguay. All procedures were performed and approved under the ethics protocols and standards required and approved by hospital authorities.

#### RNA extraction, cDNA synthesis and amplification

RNA was extracted from 140  $\mu$ l CSF samples with the QIAamp viral RNA Kit (QIAGEN) according to the manufacturer's instruction. The extracted RNA was eluted from the columns with 50  $\mu$ l RNase free water. cDNA synthesis and PCR amplification of the CVB polymerase (3D pol) were carried out as previously described [14]. Amplicons were purified using the DNA extraction Kit (Fermentas).

\*Corresponding author: Juan Cristina, Laboratorio de Virología Molecular, Centro de Investigaciones Nucleares, Facultad de Ciencias, Uruguay, Tel: +598 2525 09 01; Fax: +598 2525 08 95; E-mail: [cristina@cin.edu.uy](mailto:cristina@cin.edu.uy)

Received February 06, 2012; Accepted March 07, 2012; Published March 09, 2012

Citation: Recarey R, Moratorio G, Colina R, Cappetta M, Uriarte R, et al. (2012) Phylogenetic Analysis of Coxsackie B Viruses Reveals Genomic Plasticity and Adaptation as Studied by Codon Usage Patterns. J Medical Microbiol Diagnosis S4:001. doi:10.4172/2161-0703.S4-001

Copyright: © 2012 Recarey R, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## Sequences

DNA sequencing was performed with an ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit<sup>®</sup> and an ABI Prism 3730 Genetic Analyzer (both from Applied Biosystems, Foster City, CA, USA) at Institute Pasteur Montevideo, Uruguay. Sequences from CVB polymerase (3D pol), from positions 5907 through 6287 (according to reference strain X05690) were obtained. Complete CVB genome sequences were obtained for all available CVB strains by means of the use of ARSA at DDBJ database (available at: <http://arsa.ddbj.nig.ac.jp/html>) and EMBL database (available at: <http://www.ebi.ac.uk/embl/Access/index.html>). For strain names, accession numbers, and genotypes, see Supplementary Information Table 1.

## Sequence alignment

Sequences were aligned using the MUSCLE program [15].

## Phylogenetic analysis

In order to assign CVB strains isolated from Uruguayan patients, a phylogenetic analysis was performed. For this purpose, the FindModel program [16] was used to identify the optimal evolutionary model that best fitted our sequence datasets. Akaike Information Criteria and the hierarchical likelihood ratio test indicated that the GTR+ $\Gamma$  model was the best fit to the dataset. Maximum likelihood phylogenetic trees were constructed under the GTR+ $\Gamma$  model using software from the PhyML program [17]. As a measure of the robustness of each node, we used an approximate Likelihood Ratio Test (aLRT), which demonstrates that the branch studied provides a significant likelihood against the null hypothesis that involves collapsing that branch of the phylogenetic tree but leaving the rest of the tree topology identical [18]. aLRT was calculated using a Shimodaira-Hasegawa-like procedure (SH-like) [19,20]. All Uruguayan strains were assigned to genotype B4 (see Supplementary Information Figure 1).

## Codon usage analysis

In order to investigate the extent of codon usage bias in CVB, we constructed two different datasets, one composed of partial 3D pol sequences and another composed by all available CVB complete genome sequences. The effective number of codons (ENC) and the frequency of use of GC<sub>3</sub>S (G+C at synonymous variable third position codons, excluding Met, Trp, and termination codons) were calculated by the use of the program Codon W (available at the Mobile server (<http://mobylye.pasteur.fr>)). ENC was used to quantify the codon usage bias of an ORF [21], which is one of the best overall estimators of absolute synonymous codon usage bias. The ENC values range from 20 to 61. The larger the extent of codon bias in a gene, the smaller the ENC value is. In an extremely biased gene where only one codon is used for each amino acid, this value would be 20; in an unbiased gene, it would be 61 [22]. The relative frequencies of dinucleotides were also calculated using this program.

## Correspondence analysis

COA is an ordination technique that identifies the major trends in the variation of the data and distributes genes along continuous axis in accordance with these trends. COA creates a series of orthogonal axis to identify trends that explain the data variation, with each subsequent axis explaining a decreasing amount of the total variation across the sample [23]. Each ORF is represented in a 59-dimensional space and

each dimension is related to the relative synonymous codon usage (RSCU) value of each triplet (excluding AUG, UGG and stop codons). This was also done using with CodonW.

## Statistical analysis

Correlation analysis was carried out using Spearman's rank correlation analysis method [24] (available at: [www.wessa.net](http://www.wessa.net)).

## Nucleotide sequence accession numbers

The nucleotide sequences obtained for the 3D pol region of CVB strains isolated in Uruguay were deposited in the EMBL Database under accession numbers FR865902 to FR865910.

## Results

### Base compositional constraints on CVB codon usage bias

An ENC-GC<sub>3</sub>S plot (the effective number of codons, ENC, plotted against G+C content at the synonymous third positions, GC<sub>3</sub>S) can be used as a strategy to investigate patterns of synonymous codon usage. Genes, whose codon choices are constrained only by a G+C mutation bias, will lie on or just below the curve of predicted values if codon usage is only due to mutational bias [21]. To investigate whether the G+C compositional variation may determine differences in codon usage among CVB strains, ENC and GC<sub>3</sub>S values were calculated and plotted for 3D pol sequences from all CVB strains enrolled in these studies, which accounts for 5,080 codons (for strain included in these studies, see Supplementary Information Table 1). The results of these studies are shown in figure 1. Mean values for ENC of  $56.36 \pm 4.63$  were obtained. Since all ENC values for CVB genes are  $> 40$ , the results found suggests that the extent of codon usage bias in CVB may be low. When the GC<sub>3</sub>S values were calculated and the ENC-GC<sub>3</sub>S plots constructed all spots lie just on or below the expected curve, indicating that the codon usage bias may be influenced by the G+C compositional constraints (see Figure 1).

When the same studies were repeated using other dataset, exclusively composed by complete genomic sequences of all available CVB strains, which accounts for 61,171 codons, roughly the same results are found (for strains included in these studies, see Supplementary Information Table 1). A mean value for ENC of  $54.54 \pm 0.55$  was obtained and all spots lie below the expected curve (see Figure 1). Moreover, since all

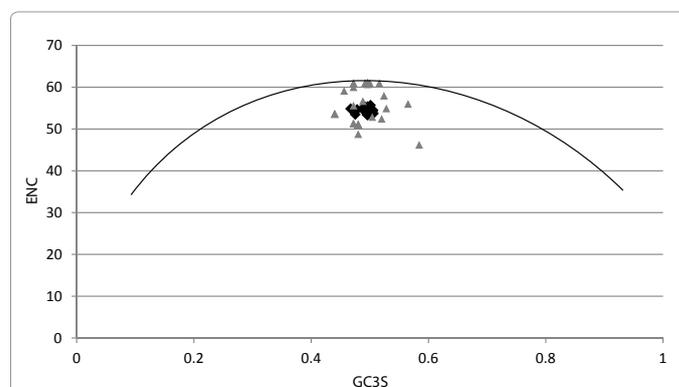


Figure 1: Effective number of codons used in each ORF plotted against the GC<sub>3</sub>S in complete CVB codes. The curve plots the relationship between GC<sub>3</sub>S and ENC in absence of selection. Gray triangle dots show the results obtained using 3D pol dataset, black square dots show the results obtained using complete CVB codes.

ENC values are  $> 50$ , we can again conclude that the codon usage bias in CVB is probably very slight.

Since codon usage by its very nature is multivariate, it is necessary to analyze the data using multivariate statistical techniques (i.e. Correspondence Analysis, COA) in order to confirm these findings. Moreover, COA has the advantage that it does not assume that the data falls into discrete clusters and therefore can represent continuous variation accurately [23]. The correlation between the position on the first axis generated by COA for each gene and the respective  $GC_3S$  values of each strain was analyzed using the complete CVB codes dataset. We have found that the position of the sequences on the first axis from COA are not correlated with the  $GC_3S$  values ( $r = -0.030$ ,  $P = 0.865$ , respectively).

Taking all together, these results suggest that factors other than gene composition contribute to codon usage among CVB strains.

### Codon usage variation among different CVB genomes

In order to detect the possibility of codon usage variation in different CVB genomes, the complete code ORFs of all available CVB strains and enrolled in these studies were divided according to their genotype (CVB 1 to 6, see Supplementary Information Table 1). A COA was performed on the RSCU values of each CVB ORF and the distribution of the six genotypes along the first two principal axes generated by the analysis was determined (Figure 2).

Surprisingly, the distribution of the six genetic groups in the plane defined by the first two major axes showed that different genotypes were located at different places, suggesting that different CVB genotypes exhibit differences in their codon usage patterns. Moreover, CVB strains isolated from different cell types and strains previously known to be associated with different diseases manifestations are also located in different sectors of the plane (Figure 2).

### Dinucleotides frequencies and codon usage in CVB genomes

It has been suggested that dinucleotide frequencies can affect codon

biases [25]. To study the possible effect of dinucleotide frequencies in CVB codon usage, the relative abundances of the 16 dinucleotides in the ORFs of the CVB strains were determined for both partial 3D pol and complete genomes (Table 1).

Roughly similar frequencies are obtained using partial 3D pol or complete CVB genomes. Interestingly, the relative abundance of CpG showed a strong deviation from the “normal range” (mean  $\pm$  S.D. =  $0.526 \pm 0.121$  and  $0.484 \pm 0.020$  for partial 3D pol and complete CVB codes, respectively) and was markedly underrepresented. On the other hand, the frequency of CpA was above the expected value (mean  $\pm$  S.D. =  $1.472 \pm 0.092$  and  $1.429 \pm 0.031$  partial 3D pol and complete CVB codes, respectively) (Table 1). Moreover, when complete CVB codes are used, 10 out of the 16 dinucleotides frequencies are highly correlated with the first axis in COA (Table 2). These observations indicate that the composition of dinucleotides plays a role in the variation found in synonymous codon usage among CVB ORFs.

Besides, the position of each codon in each of the four major axes of COA was determined for complete CVB ORFs. Table 3 shows the codons for which the maximum and minimum values were obtained for each of the axes studied (i.e. the most divergent codons values), indicating bias in their use by CVB. As it can be seen in the table, most of the divergent codons were triplets coding for Arg and Leu.

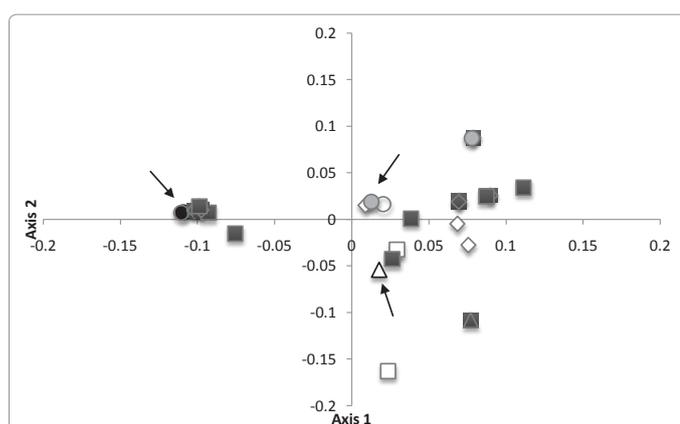
### Discussion

The results of these studies revealed that the 3 D pol region of the CVB genome contains robust phylogenetic information, in agreement with previous studies [26]. Partial 3D sequences are useful for rapid genotype assignment (Supplementary Information Figure 1). Comparable results were found using partial 3D pol region or full, complete genome sequences of CVB (Figure 1 and Table 1).

The results of this work are in agreement with previous results found for other viruses such as H5N1 Influenza A Virus (mean ENC = 50.91) [27,28]; SARS (mean ENC = 48.99) [29]; foot-and-mouth disease virus (mean ENC = 51.42) [30]; classical swine fever virus (mean ENC = 51.7) [26] and Duck Enteritis virus (mean ENC = 52.17) [31]. The ENC values found for CVB are comparatively similar to the ones found for these viruses (mean ENC value of 56.36), indicating that the overall extent of codon usage bias in CVB is not significant.

No significant correlations between the first axis of COA and  $GC_3S$  values were obtained using complete CVB genomes. This result suggest that although mutational pressure is present (Figure 1) does not significantly contribute to the codon usage bias in CVB strains, suggesting that other additional factors may be also contributing to the codon usage bias observed in CVB strains [13].

Unexpectedly, the distribution of the six genetic groups in the plane generated by the two major axes of COA showed that different genotypes are distantly located (Figure 2). Since species with a close genetic relationship always present a similar codon usage pattern [32], these findings suggest that codon usage in CVB is undergoing an evolutionary process, which is probably the result of natural selection to re-adapt its codon usage to different environments (Figure 2). This is particularly important since all six types of CVB bind and enter cells through a common receptor protein (CAR) [8,9], present in different cell types that CVB can infect. Indeed, CVB is known to be associated with different cell types and disease manifestations are distantly located in the plane, suggesting a codon usage adaptation to the infected cell (Figure 2). This is also in agreement with COA analysis, where most of



**Figure 2: Positions of the 28 complete ORFs of CVB 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 32.31% and 12.52% of the total variation, respectively. The CVB ORFs are divided according to their genotype, genotype B1 strains are indicated by a black diamond (◆), genotype B2 by a white square (□), genotype B3 by a black square (■), genotype B4 by a white diamond (◇), genotype B5 by a black triangle (▲) and genotype B6 by gray circle. CVB strains known to be associated with diabetes (DQ480420, genotype B4) or myocarditis (M16572, genotype B3) are shown by a black (●) and a white (○) circle, respectively. Strain AY875692 (genotype B5), isolated from a CSF sample, is shown by a white triangle (Δ).

Partial 3D pol								
	UU	UC	UA	UG	CU	CC	CA	CG
Mean ± SD <sup>a</sup>	0.899±0.092	0.746 ± 0.089	0.798 ± 0.096	0.986 ± 0.116	0.757 ± 0.083	0.770 ± 0.110	1.472 ± 0.092	0.526 ± 0.121
	AU	AC	AA	AG	GU	GC	GA	GG
Mean ± SD	0.980 ± 0.108	1.147 ± 0.110	1.482 ± 0.128	1.339 ± 0.095	0.794 ± 0.139	0.865 ± 0.118	1.258 ± 0.092	1.116 ± 0.095
Complete codes								
	UU	UC	UA	UG	CU	CC	CA	CG
Mean ± SD	0.900 ± 0.042	0.758 ± 0.024	0.844 ± 0.049	1.238 ± 0.047	0.850 ± 0.023	0.909 ± 0.030	1.429 ± 0.031	0.484 ± 0.020
	AU	AC	AA	AG	GU	GC	GA	GG
Mean ± SD	1.038 ± 0.031	1.144 ± 0.044	1.283 ± 0.028	1.169 ± 0.029	0.954 ± 0.019	0.857 ± 0.028	1.078 ± 0.018	1.057 ± 0.028

a Mean values of relative dinucleotide ratios ± standard deviation.

**Table 1:** Relative abundance of dinucleotides in CVB ORFs.

		UU	UC	UA	UG	CU	CC	CA	CG
Axis 1	<i>r</i>	0.660	0.680	-0.779	0.774	0.311	-0.711	-0.035	-0.370
	<i>P</i>	<0.001	<0.001	<0.001	<0.001	0.105	<0.001	0.849	0.053
		AU	AC	AA	AG	GU	GC	GA	GG
Axis 1	<i>r</i>	0.579	-0.621	0.683	-0.715	-0.396	0.730	-0.157	-0.260
	<i>P</i>	<0.01	0.001	<0.001	<0.001	0.038	<0.001	0.406	0.173

**Table 2:** Summary of correlation analysis between the first axis in COA and sixteen dinucleotides frequencies in complete CVB codes.

Axis 1			Axis 2		
Codon	Value	Aminoacid	Codon	Value	Aminoacid
CGG	-0.334	Arg	UCG	-0.143	Ser
CGA	0.289	Arg	GUU	0.082	Val
Axis 3			Axis 4		
Codon	Value	Aminoacid	Codon	Value	Aminoacid
CUU	0.065	Leu	UUA	-0.142	Leu
CUC	0.096	Leu	CUU	0.097	Leu

**Table 3:** Position of codons in each of the four major axes of COA in CVB ORFs.

the divergent codons were triplets coding for Arg and Leu (Table 3). This reveals that the use of Arg and Leu codons play also a role in the evolution and the variability observed among CVB strains.

CpG containing codons were markedly underrepresented in CVB ORF's (Table 1). This result is in agreement with previous results found for Coronaviruses [33], vertebrate-infecting members of the family *Flaviviridae* [34], Poliovirus [35] and Hepatitis A virus [36]. CpG deficiency has been related to the immunostimulatory properties of unmethylated CpG, which are recognized by the host's innate immune system as a pathogen signature [13,32]. Since the vertebrate immune system relies on unmethylated CpG recognition in DNA molecules as a sign of infection, the under-representation in RNA viruses of this dinucleotide, has been so far exclusively observed in vertebrate viruses [33]. Thus, escaping from the host antiviral response may act as another selective pressure contributing to the multifactorial codon usage shaping of CVB strains [37].

The results of these studies suggest that CVB genomic biases are the result of co-evolution of translation adaptation and probably the need to escape anti-viral cell defenses. This is in agreement with the evolution rhetoric theory proposed by Vetsigian & Goldenfeld [37] in which genome biases emerge by the need to increase communication with the ever changing cell environment without changing the message.

#### Acknowledgements

The authors acknowledge support from Agencia Nacional de Investigación e Innovación (ANII) through project PE\_ALI\_2009\_1\_1603 and PEDECIBA, Uruguay.

#### References

1. Khetsuriani N, Lamonte-Fowlkes A, Oberst S, Pallansch MA Centers for Disease Control and Prevention (2006) Enterovirus surveillance--United States, 1970-2005. *MMWR Surveill Summ* 55: 1-20.
2. Castro CM, Oliveira DS, Macedo O, Lima MJ, Santana MB, et al. (2009) Echovirus 30 associated with cases of aseptic meningitis in state of Pará, Northern Brazil. *Mem Inst Oswaldo Cruz* 104: 444-450.
3. Fariás A, Cabrerizo M, Ré V, Glatstein N, Pisano B, et al. (2011) Molecular identification of human enteroviruses in children with neurological infections from the central region of Argentina. *Arch Virol* 156: 129-133.
4. Pinto Junior VL, Rebelo MC, Costa EV, Silva EE, Bóia MN (2009) Description of a widespread outbreak of aseptic meningitis due to echovirus 30 in Rio de Janeiro state, Brazil. *Braz J Infect Dis* 13: 367-370.
5. Yoon SY, Ha YE, Choi JE, Ahn J, Lee H, et al. (2008) Coxsackievirus B4 uses autophagy for replication after calpain activation in rat primary neurons. *J Virol* 82: 11976-11978.
6. Wang JP, Cerny A, Asher DR, Kurt-Jones EA, Bronson RT, et al. (2010) MDA5 and MAVS mediate type I interferon responses to coxsackie B virus. *J Virol* 84: 254-260.
7. Hyöty H, Taylor KW (2002) The role of viruses in human diabetes. *Diabetologia* 45: 1353-1361.
8. Zanone MM, Favaro E, Ferioli E, Huang GC, Klein NJ, et al. (2007) Human pancreatic islet endothelial cells express coxsackievirus and adenovirus receptor and are activated by coxsackie B virus infection. *FASEB J* 21: 3308-3317.
9. Bergelson JM, Cunningham JA, Droguett G, Kurt-Jones EA, Krithivas A, et al. (1997) Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* 275: 1320-1323.
10. Duret L (2002) Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev* 12: 640-649.

11. Jenkins GM, Holmes EC (2003) The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res* 92: 1-7.
12. Kliman RM, Bernal CA (2005) Unusual usage of AGG and TTG codons in humans and their viruses. *Gene* 352: 92-99.
13. Shackelton LA, Parrish CR, Holmes EC (2006) Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J Mol Evol* 62: 551-563.
14. Lindberg AM, Andersson P, Savolainen C, Mulders MN, Hovi T (2003) Evolution of the genome of Human enterovirus B: incongruence between phylogenies of the VP1 and 3CD regions indicates frequent recombination within the species. *J Gen Virol* 84: 1223-1235.
15. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
16. Posada D, Crandall KA (2001) Selecting the best-fit model of nucleotide substitution. *Syst Biol* 50: 580-601.
17. Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* 33: W557-559.
18. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 55: 539-552.
19. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51: 492-508.
20. Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246-1247.
21. Wright F (1990) The 'effective number of codons' used in a gene. *Gene* 87: 23-29.
22. Comeron JM, Aguadé M (1998) An evaluation of measures of synonymous codon usage bias. *J Mol Evol* 47: 268-274.
23. Greenacre MJ (1984) Theory and applications of correspondence analysis. Academic Press, London.
24. Wessa P (2010) Free Statistics Software, Office for Research Development and Education, version 1.1.23-r5.
25. Bolanaki E, Kottaridi C, Markoulatos P, Kyriakopoulou Z, Margaritis L, et al. (2007) Partial 3D gene sequences of Coxsackie viruses reveal interspecies exchanges. *Virus Genes* 35: 129-140.
26. Ahn I, Jeong BJ, Bae SE, Jung J, Son HS (2006) Genomic analysis of influenza A viruses, including avian flu (H5N1) strains. *Eur J Epidemiol* 21: 511-519.
27. Tao P, Dai L, Luo M, Tang F, Tien P, et al. (2009) Analysis of synonymous codon usage in classical swine fever virus. *Virus Genes* 38: 104-112.
28. Zhou T, Gu W, Ma J, Sun X, Lu Z (2005) Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. *Biosystems* 81: 77-86.
29. Zhao S, Zhang Q, Liu X, Wang X, Zhang H, et al. (2008) Analysis of synonymous codon usage in 11 human bocavirus isolates. *Biosystems* 92: 207-214.
30. Zhong J, Li Y, Zhao S, Liu S, Zhang Z (2007) Mutation pressure shapes codon usage in the GC-Rich genome of foot-and-mouth disease virus. *Virus Genes* 35: 767-776.
31. Jia R, Cheng A, Wang M, Xin H, Guo Y, et al. (2009) Analysis of synonymous codon usage in the UL24 gene of duck enteritis virus. *Virus Genes* 38: 96-103.
32. Sharp PM, Cowe E, Higgins DG, Shields DC, Wolfe KH, et al. (1988) Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Res* 16: 8207-8211.
33. Woo PCY, Wong BH, Huang Y, Lau SK, Yuen KY (2007) Cytosine deamination and selection of CpG suppressed clones are the two major independent biological forces that shape codon usage bias in Coronaviruses. *Virology* 369: 431-442.
34. Lobo FP, Mota BE, Pena SD, Azevedo V, Macedo AM, et al. (2009) Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. *PLoS One* 4: e6282.
35. Rothberg PG, Wimmer E (1981) Mononucleotide and dinucleotide frequencies, and codon usage in poliovirus RNA. *Nucleic Acids Res* 9: 6221-6229.
36. Bosch A, Mueller S, Pinto RM (2010) Coding biases and viral fitness. In: Ehrenfeld E, Domingo E, Roos RP (eds) *The picornaviruses*. ASM Press, Washington, DC. pp 271-284.
37. Vetsigian K, Goldenfeld N (2009) Genome rhetoric and the emergence of compositional bias. *Proc Natl Acad Sci U S A* 106: 215-220.

This article was originally published in a special issue, [Molecular Mechanisms of Viruses](#) handled by Editor(s). Dr. Joel Kenneth Weltmann, Alpert Brown University, USA; Dr. Sumathi Sankaran, University of California, USA