



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Research paper

The adaptation of codon usage of +ssRNA viruses to their hosts

Lin Tian^{a,b}, Xuejuan Shen^a, Robert W. Murphy^c, Yongyi Shen^{a,b,d,*}^a College of Veterinary Medicine, South China Agricultural University, Guangzhou 510642, China^b Shantou University Medical College, Shantou 515041, China^c Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, Toronto M5S 2C6, Canada^d Key Laboratory of Zoonosis Prevention and Control of Guangdong Province, Guangzhou 510642, China

ARTICLE INFO

Keywords:

Codon usage patterns
tRNA pool
Adaptability
Expression system
Virus
Host

ABSTRACT

Viruses depend on their host's cellular structure to survive. Most of them do not have tRNAs, their translation relies on hosts' tRNA pools. Over the course of evolution, viruses needed to optimally exploit cellular processes of their host. Thus, codon usage of a virus should coevolve with its host to efficiently and rapidly replicate. Some viruses can invade a broad spectrum of hosts (BSTVs), while others can invade a narrow spectrum only (NSTVs). Consequently, we test the hypothesis that similarity of codon usage preference and the degree of matching between BSTVs and their hosts will be lower than that of NSTVs, which only need to coevolve with few hosts. We compare the patterns of codon usage in 255 virus genomes to test this hypothesis. Our results show that NSTVs have a higher degree of matching to their hosts' tRNA pools than BSTVs. Further, analysis of the effective number of codons (ENC) infers that codon usage bias of NSTVs is relatively stronger than that of BSTVs. Thus, codon usage of NSTVs tends to better match their host than that of BSTVs. This supports the hypothesis that viruses adapt to the expression system of their host(s).

1. Introduction

Viruses are pure parasites. They depend on their hosts' cellular structure and metabolism to replicate and assemble, i.e., survive. Most of their genomes do not encode tRNAs, thus their translation of viral proteins relies on the hosts' tRNA pools (Kumar et al. 2016). A successful infection requires that viruses possess the ability to enter the host cell, and efficiently produce new viruses. The degenerate genetic code unequally uses synonymous codons, which code for the same amino acid (Cristina et al. 2016; Kanaya et al. 2001; Shackelton et al. 2006; Tsai et al. 2007). The redundancy of the genetic code provides the opportunity to shape the efficiency and accuracy of protein production, while maintaining the same amino acid sequence (Chaney and Clark 2015; Plotkin and Kudla 2011; Stoletski and Eyre-Walker 2007). Considering that the translation of viral proteins relies on the host's pool of tRNAs, codon usage of a virus must coevolve with its host to efficiently use host resources. It is expected that higher similarity of codon usage pattern will better facilitate their replication.

The extent of codon usage among viruses and their hosts has been suggested to affect viral survival, fitness, and evasion from host's immune system (Burns et al. 2006; Costafreda et al. 2014; Mueller et al. 2006). Because the virus relies on the host's cellular machinery for its replication, codon usage bias was suggested to play a role in the

adaptation of a virus to its host. Codon usage bias is common in viruses (Butt et al. 2014; Castells et al. 2017; Cristina et al. 2016; He et al. 2017; Li et al. 2017; Moratorio et al. 2013; Singh et al. 2016; Su et al. 2017; Xu et al. 2017; Zang et al. 2017; Zhao et al. 2016). Efficient replication seemingly requires that a virus and host have similar codon usage patterns to share a tRNA pool. Co-evolution between a certain RNA virus and its susceptible hosts at codon usages have been observed in many viruses (Franzo et al. 2017; Rahman et al. 2017; Simón et al. 2017).

Some viruses have a broad ranges of hosts (BSTVs), such as arbovirus. These can infect mammals, birds, and insects. Other viruses have narrow host ranges (NSTVs), which can infect a limited number of hosts only. Because BSTVs must fit to multiple hosts and their diverse tRNA pools and their codon usage has a relationship with their host, a tradeoff exists regarding the extent of codon usage. BSTVs must fit to their diverse hosts and, thus, the extent of matching for codon usage would be lower than that of NSTVs, which must fit few, similar hosts. To test this hypothesis, we analyze 255 viruses from 20 genera of positive-sense single-stranded RNA (+ssRNA) viruses (e.g., *Flavivirus*, *Alphavirus*, *Coronavirus*, *Torovirus*, *Arterivirus*, *Rubivirus*, *Pestivirus*, *Hepacivirus*, *Alphamesonivirus*). Viruses that can infect vertebrates and invertebrates, such as most of the *Flavivirus* and *Alphavirus*, were classified as BSTVs, while other viruses that can infect either vertebrate or

* Corresponding author at: College of Veterinary Medicine, South China Agricultural University, Guangzhou 510642, China.
E-mail address: sheny@scau.edu.cn (Y. Shen).

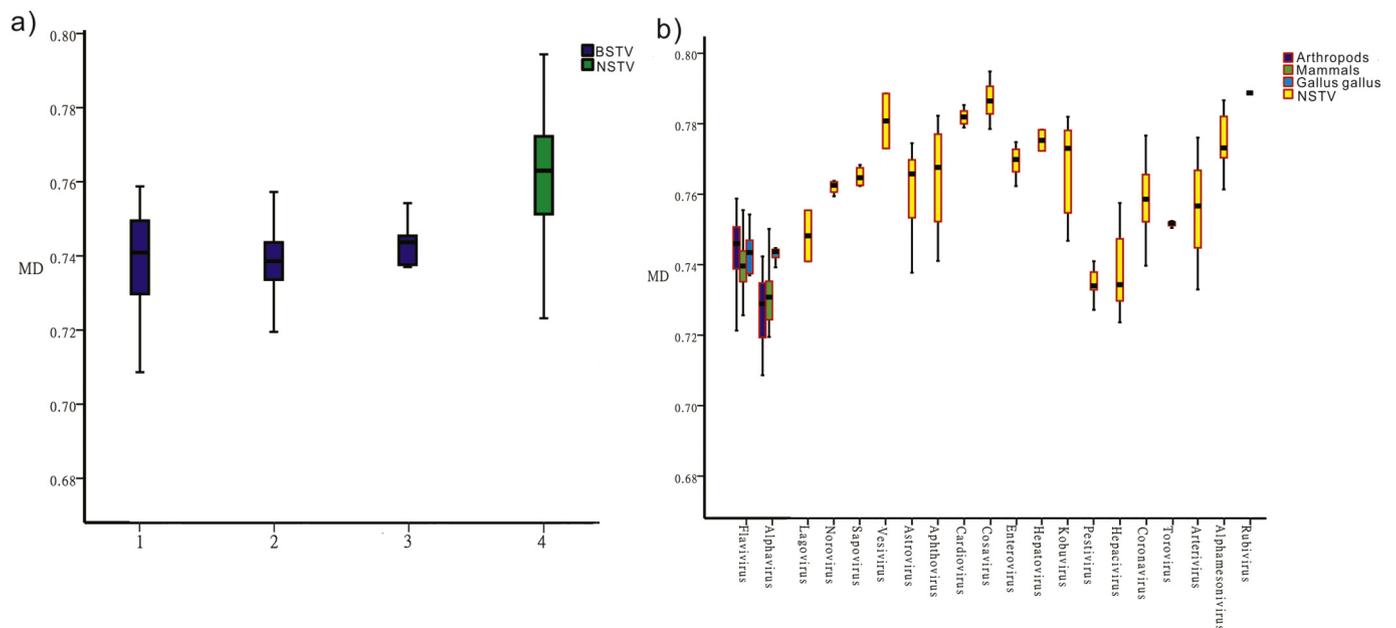


Fig. 1. Matching degree (MD) of viruses to their hosts' tRNA pools.

(a) Group 1 represents MD of BSTVs to arthropod's tRNA pool. Group 2 shows the MD of BSTVs to mammal's tRNA pool. Group 3 is the MD of BSTVs to *Gallus gallus*'s tRNA pool. Group 4 shows the MD of NSTVs to their hosts' tRNA pools. (b) The MD of 20 virus genera (+ ssRNA) to their hosts' tRNA pools.

invertebrate were classified as NSTVs.

2. Materials and methods

2.1. Data collection and preliminary analyses

The complete genome sequences of 255 virus strains from the 20 genera of + ssRNA viruses were obtained from the GenBank database (<http://www.ncbi.nlm.nih.gov>). The information of host range was determined from NCBI (<https://www.ncbi.nlm.nih.gov/taxonomy/>) and the Ninth Report of the International Committee on Taxonomy of Viruses (ViralZone Database: <http://viralzone.expasy.org/>). Accession numbers and other detailed information of these viruses, such as strain names, isolate hosts and host ranges were also retrieved (Supplementary Table 1).

Estimates of codon usage, the relative synonymous codon usage (RSCU) (Sharp and Li 1986), and the effective number of codons (ENC) (Wright 1990), were calculated using CodonW (available at <http://sourceforge.net/projects/codonw>). In this study, *Aedes*, *Culex* and *Ixodes* represented the main arthropod hosts, and *Gallus*, *Homo* and *Mus* the three major groups of vertebrates. Coding sequences of the hosts were obtained from the Ensembl database (available at: <http://www.ensembl.org>) (Yates et al. 2016). Copy number of tRNAs for transmission vectors and hosts were obtained from GtRNAdb (<http://gttrnadb.ucsc.edu/>).

2.2. Estimating the adaptability of viruses to their hosts' tRNA pools

Optimizing codon usage of viruses according to that of highly expressed host genes has been proved to increase the production of viral proteins (Chithambaram et al. 2014; Ngumbela et al. 2008) or transgenic genes (Koresawa et al. 2000). The degree of similarity for overall codon usage between viruses and their hosts' tRNA pools was estimated with a parameter based on optimized codon usage and the extent of matching between viral ORF's codon usage bias and their hosts' tRNA pools. ORFs were optimized on the basis of tRNA copy number characteristics of their hosts' expression system. Online optimization software (<http://genomes.urv.es/OPTIMIZER/>) (Puigbo et al. 2007) was utilized. The matching degree (MD) was calculated as follows:

$$MD = \frac{M}{N}$$

where M was defined as the number of the different bases before and after optimized sequence, and N was the total length of open reading frame. This value could have ranged from zero to 1.

To better quantify the effect of the overall codon usage of the host on the formation of the overall codon usage of the virus, the similarity index $D(A,B)$ reported by the previous study was introduced into our work (Zhou et al. 2013). The $D(A,B)$ represented the potential effect of the overall codon usage of the host on that of virus. This value potentially ranged from zero to 1.0.

3. Results

3.1. The matching degree of + ssRNA viruses to their hosts' tRNA pools

MD values were calculated for viral ORFs. Unfortunately, some viruses lacked data of tRNA copy numbers, and coding sequence of their hosts. Therefore, only 255 viruses (+ ssRNA) were used in this analysis. The 101 strains of arboviruses that belonged to BSTVs were optimized according to their hosts' tRNA pool expression systems (host: arthropods, mammals, *Gallus gallus*). MD values (mean \pm SD) were 0.7388 ± 0.0121 , 0.7383 ± 0.0083 , and 0.7427 ± 0.0045 in the three hosts, respectively (Supplementary Table 2). The MD values of 154 NSTVs strains had a mean of 0.7617 ± 0.0168 (Supplementary Table 3). Wilcoxon & Mann-Whitney *U* test obtained statistically significant higher MD values in NSTVs than BSTVs ($Z = -9.99$, $p < 0.001$; $Z = -10.25$, $p < 0.001$; $Z = -5.59$, $p < 0.001$, respectively. Fig. 1a).

Among the different genera of arbovirus (Fig. 1b), the extent of matching for *Flavivirus* to their hosts' tRNA pools was higher than that of *Alphavirus*. In addition, among the 18 genera of NSTVs, the MD values of *Togaviridae* (*rubivirus*) were the highest, followed by *Picornaviridae* (*Aphthovirus*, *Cardiovirus*, *Casavirus*, *Enterovirus*, *Hepatovirus*, *Kobuvirus*), *Mesoniviridae* (*Alphamesonivirus*), *Caliciviridae* (*Lagovirus*, *Norovirus*, *Sapovirus*, *Vesivirus*), *Arterviridae* (*Arterivirus*), *Astroviridae* (*Astrovirus*), *Coronaviridae* (*Coronavirus*, *Torovirus*), and the last was *Flaviridae* (*Pestivirus*, *Hepacivirus*).

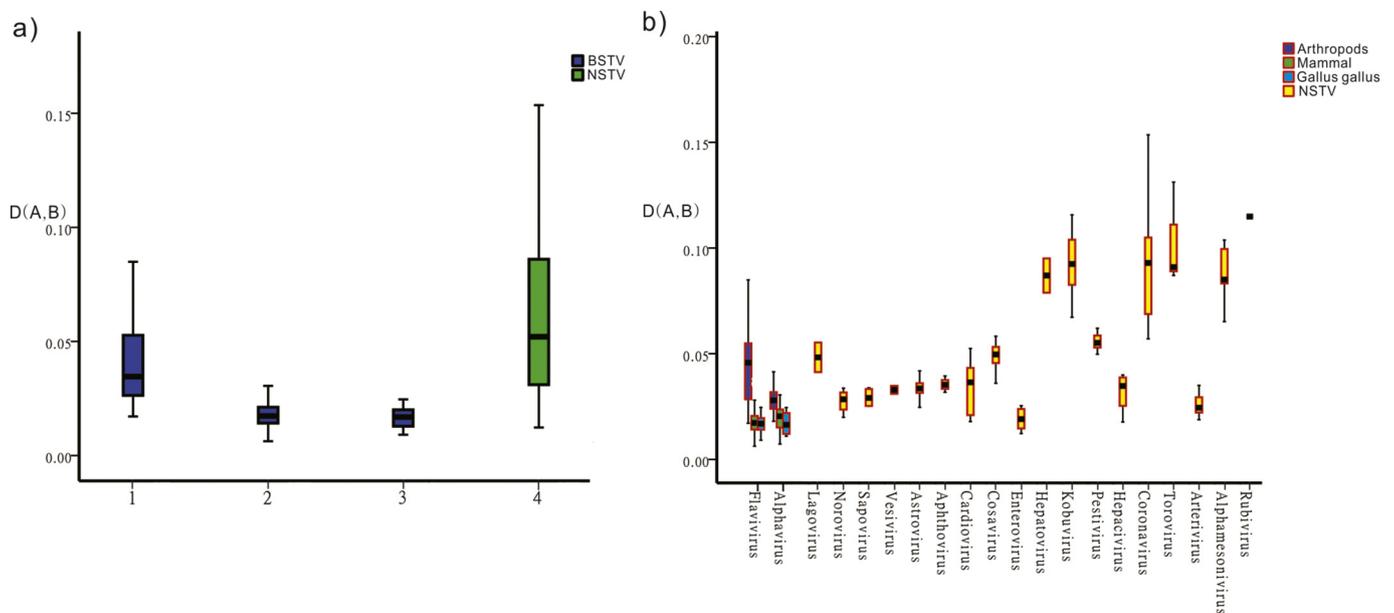


Fig. 2. Similarity of overall codon usage pattern of viruses to their hosts.

(a) Group 1 is the similarity degree between BSTVs and arthropods. Group 2 is the similarity degree between BSTVs and mammals. Group 3 is the similarity degree between BSTVs and *Gallus gallus*. Group 4 is the similarity degree between NSTVs and a particular host. (b) The similarity degree of the overall codon usage bias between 20 virus genera (+ ssRNA) and the hosts.

3.2. Similarity of overall codon usage between +ssRNA viruses and their hosts

Similarity index values $D(A,B)$ were obtained for each strain in relation to its host(s) (Supplementary Tables 2 and 3). As shown in Fig. 2a, the indices of the fourth group (NSTVs vs hosts) were higher than those of the groups 1, 2, and 3 (BSTVs vs arthropods, BSTVs vs mammals, BSTVs vs *Gallus gallus*).

To quantify the degree of similarity of the overall codon usage pattern between 20 different virus genera and their hosts, the similarity index $D(A,B)$ was calculated to all strains (Fig. 2b). The MD and $D(A,B)$ values showed a similar trend in that values of NSTVs were the highest.

Finally ENC's values were calculated for the 255 strains to gain insights into codon usage bias. All viruses had ENC values > 40, except Human coronavirus HKU1 (ENC = 35.66), Human coronavirus NL63 (ENC = 36.77), Rubella virus (ENC = 38.45), Human hepatitis A virus (ENC = 39.23), Canine kobuvirus 1 (ENC = 39.38), Aichi virus (ENC = 39.52), Simian hepatitis A virus (ENC = 39.91), and Kobuvirus dog/AN211D/USA/2009 (ENC = 39.92). The codon preferences of NSTVs were stronger than for the BSTVs (Fig. 3).

4. Discussion

The degeneracy of the genetic code implies that several triplets can code for the same amino acid. The use of synonymous codons in gene coding regions does not occur randomly, and codon usage bias is very common in viruses (Butt et al. 2014; Cristina et al. 2016; He et al. 2017; Moratorio et al. 2013; Singh et al. 2016). Codon usage is among the determinant factors that influence gene expression levels (Chaney and Clark 2015; Zhou et al. 2016). Because viruses do not have tRNAs, and rely on host cell machinery for replication, co-evolution between a certain RNA virus and its susceptible hosts at codon usages have been observed (Franzo et al. 2017; Rahman et al. 2017; Simón et al. 2017). However, ambiguity remains in the co-evolution patterns of different viruses. Some viruses infect a broad range of species (BSTVs), whereas others infect only a single host (NSTVs). Viruses have very diverse hosts, and different hosts have very diverse tRNA pools. The MD and $D(A,B)$ values of NSTVs are significantly higher than those of the BSTVs

vs. *Anopheles gambiae*, *Homo sapiens*, *Gallus gallus*, and *Macaca mulatta* (Figs. 1 and 2). Thus, and as our hypothesis predicts, NSTVs appear to be more precisely adapted to their hosts' codon usage pattern and tRNA pools than BSTVs. Each NSTV can infect one host only. Therefore, these viruses need to fit only one host's tRNA pool. They appear to have evolved more consistent codon usage patterns with their hosts' expression systems. In contrast, BSTVs can infect and replicate in mammals, birds, and insects. Thus, adaptation in BSTVs may involve a tradeoff between precise and functional matching to fit the diverse tRNA pools of multiple hosts. This may explain the relatively lower matching of BSTVs to their hosts.

Enterovirus, *Hepacivirus* and *Arterivirus* are NSTVs and their $D(A,B)$ values are relatively low (Fig. 2b). This indicates a relatively low extent of similarity in overall codon usage between these viruses and their host. These viruses may not need to replicate rapidly. Other factors, such as mutational pressure, may also play a role in determining codon usage bias (Gu et al. 2004; Rahman et al. 2017; Wang et al. 2011; Wong et al. 2010).

To quantify the extent of variation in codon usage, the ENC values were calculated (Wright 1990). Most viruses have ENC values > 40, which represents weak codon bias. This may be beneficial for efficient replication of viruses in host cells with potentially distinct codon preferences. The codon preference of NSTVs is relatively stronger than for BSTVs (Fig. 3). This is could be due to weak codon preference being advantageous in the adaptation of BSTVs to multiple host expression systems.

The ability to enter the host-cell and efficiently replicate itself is essential for viral infection. Viruses have coevolved many pathways to transcribe their own genetic material in their hosts (Harwig et al. 2017). Codon usage in BSTVs may involve a tradeoff between precise and functional matching to fit the diverse tRNA pools of multiple hosts. As expected, our analysis show that generally NSTVs are more adapted to their hosts' codon usage pattern and tRNA pools than BSTVs. This may help the virus to use the host transcript machinery more efficiently and, therefore, replicate faster.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.meegid.2018.05.034>.

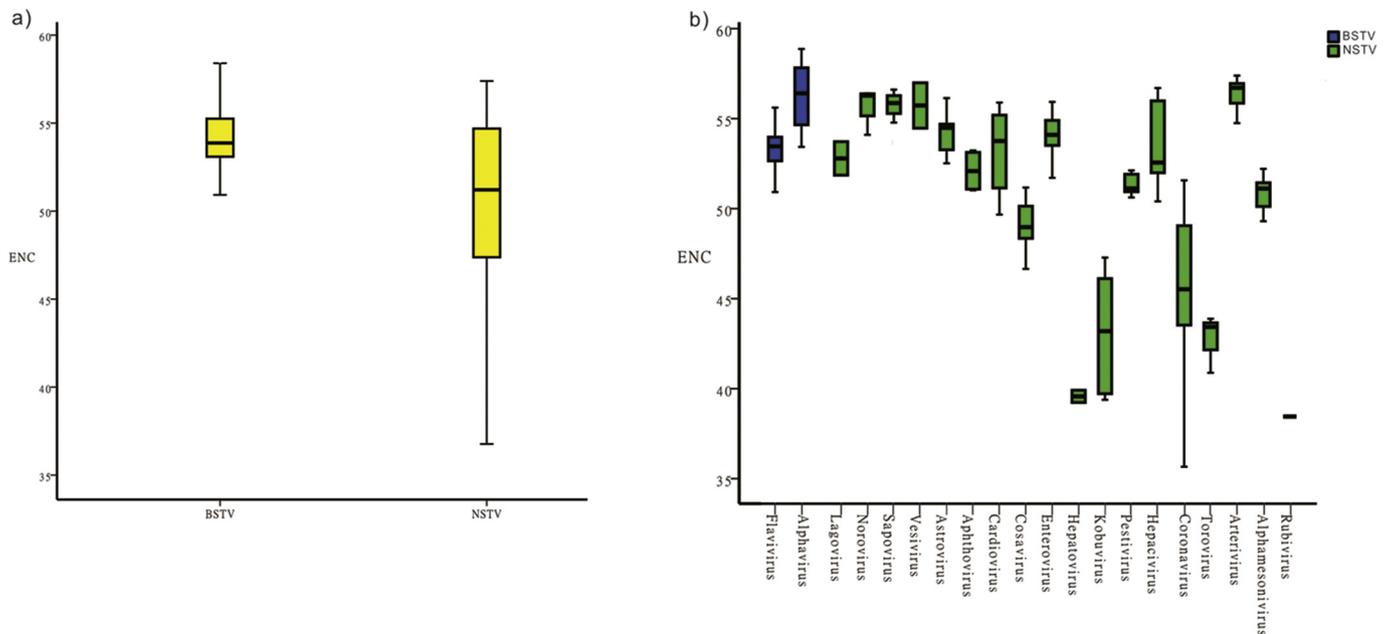


Fig. 3. Effective number of codons (ENCs) of 20 virus genera.

(a) BSTVs and NSTVs. (b) 20 virus genera (+ssRNA).

Author contributions

YS: designed the study; LT and XS: analyzed the data; LT, RWM and YS: drafted the manuscript; All authors read and approved the final manuscript.

Acknowledgments

This work was supported by Guangdong Natural Science Funds for Distinguished Young Scholar (2014A030306046).

References

- 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.
- Castells, M., Victoria, M., Colina, R., Musto, H., Cristina, J., 2017. Genome-wide analysis of codon usage bias in *Bovine Coronavirus*. *Virology* 14, 115.
- Chaney, J.L., Clark, P.L., 2015. Roles for synonymous codon usage in protein biogenesis. *Annu Rev Biophys* 44, 143–166.
- Chithambaram, S., Prabhakaran, R., Xia, X., 2014. The effect of mutation and selection on codon adaptation in *Escherichia coli* bacteriophage. *Genetics* 197, 301–315.
- Costafreda, M.I., Perez-Rodriguez, F.J., D'Andrea, L., Guix, S., Ribes, E., Bosch, A., Pinto, 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.
- Cristina, J., Fajardo, A., Sonora, M., Moratorio, G., Musto, H., 2016. A detailed comparative analysis of codon usage bias in Zika virus. *Virus Res* 223, 147–152.
- Franzo, G., Tucciarone, C.M., Cecchinato, M., Drigo, M., 2017. Canine parvovirus type 2 (CPV-2) and Feline panleukopenia virus (FPV) codon bias analysis reveals a progressive adaptation to the new niche after the host jump. *Mol Phylogenet Evol* 114, 82–92.
- Gu, W., Zhou, T., Ma, J., Sun, X., Lu, Z., 2004. Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. *Virus Res* 101, 155–161.
- Harwig, A., Landick, R., Berkhout, B., 2017. The battle of RNA synthesis: virus versus host. *Viruses* 9 (pii: E309).
- He, W., Zhang, H., Zhang, Y., Wang, R., Lu, S., Ji, Y., Liu, C., Yuan, P., Su, S., 2017. Codon usage bias in the N gene of rabies virus. *Infect Genet Evol* 54, 458–465.
- Kanaya, S., Yamada, Y., Kinouchi, M., Kudo, Y., Ikemura, T., 2001. Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J Mol Evol* 53, 290–298.
- Koresawa, Y., Miyagawa, S., Ikawa, M., Matsunami, K., Yamada, M., Shirakura, R., Okabe, M., 2000. Synthesis of a new Cre recombinase gene based on optimal codon usage for mammalian systems. *J Biochem* 127, 367–372.

- Kumar, N., Bera, B.C., Greenbaum, B.D., Bhatia, S., Sood, R., Selvaraj, P., Anand, T., Tripathi, B.N., Virmani, N., 2016. Revelation of influencing factors in overall codon usage bias of equine influenza viruses. *PLoS One* 11, e0154376.
- Li, G., Ji, S., Zhai, X., Zhang, Y., Liu, J., Zhu, M., Zhou, J., Su, S., 2017. Evolutionary and genetic analysis of the VP2 gene of canine parvovirus. *BMC Genomics* 18, 534.
- 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.
- Ngumbela, K.C., Ryan, K.P., Sivamurthy, R., Brockman, M.A., Gandhi, R.T., Bhardwaj, N., Kavanagh, D.G., 2008. Quantitative effect of suboptimal codon usage on translational efficiency of mRNA encoding HIV-1 gag in intact T cells. *PLoS One* 3, 0002356.
- Plotkin, J.B., Kudla, G., 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* 12, 32–42.
- Puigbo, P., Guzman, E., Romeu, A., Garcia-Vallve, S., 2007. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res* 35, W126–W131.
- Rahman, S.U., Yao, X., Li, X., Chen, D., Tao, S., 2017. Analysis of codon usage bias of Crimean-Congo hemorrhagic fever virus and its adaptation to hosts. *Infect Genet Evol* 58, 1–16.
- 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., Li, W.H., 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Res* 14, 7737–7749.
- Simón, D., Fajardo, A., Sonora, M., Delfraro, A., Musto, H., 2017. Host influence in the genomic composition of flaviviruses: a multivariate approach. *Biochem Biophys Res Commun* 492, 572–578.
- Singh, N.K., Tyagi, A., Kaur, R., Verma, R., Gupta, P.K., 2016. Characterization of codon usage pattern and influencing factors in Japanese encephalitis virus. *Virus Res* 221, 58–65.
- Stoletzki, N., Eyre-Walker, A., 2007. Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol Biol Evol* 24, 374–381.
- Su, W., Li, X., Chen, M., Dai, W., Sun, S., Wang, S., Sheng, X., Sun, S., Gao, C., Hou, A., Zhou, Y., Sun, B., Gao, F., Xiao, J., Zhang, Z., Jiang, C., 2017. Synonymous codon usage analysis of hand, foot and mouth disease viruses: a comparative study on coxsackievirus A6, A10, A16, and enterovirus 71 from 2008 to 2015. *Infect Genet Evol* 53, 212–217.
- Tsai, C.T., Lin, C.H., Chang, C.Y., 2007. Analysis of codon usage bias and base compositional constraints in iridovirus genomes. *Virus Res* 126, 196–206.
- Wang, M., Liu, Y.S., Zhou, J.H., Chen, H.T., Ma, L.N., Ding, Y.Z., Liu, W.Q., Gu, Y.X., Zhang, J., 2011. Analysis of codon usage in Newcastle disease virus. *Virus Genes* 42, 245–253.
- 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* 10, 1471–2148.
- Wright, F., 1990. The 'effective number of codons' used in a gene. *Gene* 87, 23–29.
- Xu, X., Fei, D., Han, H., Liu, H., Zhang, J., Zhou, Y., Xu, C., Wang, H., Cao, H., Zhang, H., 2017. Comparative characterization analysis of synonymous codon usage bias in classical swine fever virus. *Microb Pathog* 107, 368–371.

- Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girón, C.G., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S.J., Cunningham, F., Aken, B.L., Zerbino, D.R., Flicek, P., 2016. Ensembl 2016. *Nucleic Acids Res* 44, D710–D716.
- Zang, M., He, W., Du, F., Wu, G., Wu, B., Zhou, Z., 2017. Analysis of the codon usage of the ORF2 gene of feline calicivirus. *Infect Genet Evol* 54, 54–59.
- Zhao, Y., Zheng, H., Xu, A., Yan, D., Jiang, Z., Qi, Q., Sun, J., 2016. Analysis of codon usage bias of envelope glycoprotein genes in nuclear polyhedrosis virus (NPV) and its relation to evolution. *BMC Genomics* 17, 677.
- Zhou, J.H., Zhang, J., Sun, D.J., Ma, Q., Chen, H.T., Ma, L.N., Ding, Y.Z., Liu, Y.S., 2013. The distribution of synonymous codon choice in the translation initiation region of dengue virus. *PLoS One* 8, e77239.
- Zhou, Z., Dang, Y., Zhou, M., Li, L., Yu, C.H., Fu, J., Chen, S., Liu, Y., 2016. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc Natl Acad Sci U S A* 113, E6117–E6125.