



# Genetic features of a translation initiation system composed of IRES element, nucleotide context surrounding the initiation codon, and translation initiation region of classical swine fever virus RNA

X.-X. Ma\*, Y.-P. Feng\*, Y.-Q. Zhao, J.-L. Liu, L. Chen, P.-H. Guo, J.-Z. Guo, L.-Y. Ma and Z.-R. Ma

College of Life Science and Engineering,  
Northwest University for Nationalities, Lanzhou, China

\*These authors contributed equally to this study.

Corresponding author: Z.-R. Ma

E-mail: maxiaoxia956@163.com / mzr@xbmu.edu.cn

Genet. Mol. Res. 13 (4): 10803-10810 (2014)

Received December 18, 2013

Accepted March 6, 2014

Published December 19, 2014

DOI <http://dx.doi.org/10.4238/2014.December.19.1>

**ABSTRACT.** Nucleotide and codon usage are typically examined to investigate viral evolution. In this study, we analyzed the genetic information of 46 strains of classical swine fever virus (CSFV) RNA, nucleotide usage in the internal ribosome entry site (IRES), the nucleotide context surrounding the initiation codon, and synonymous codon usage in the translation initiation region. Phylogenetic analysis of the IRES element indicated that the genetic diversity of this element is generally similar to the phylogenetic clusters of CSFV genotypes. Nucleotides surrounding the initiation codon of CSFV RNA were generally more stable (ACAUGGCACAAUGGAGUUG) compared to the internal AUG in the CSFV coding sequence. The second codon position after the initiation codon was generally selected to be GAG, which has

lower tRNA abundance in pigs than its synonymous member (GAA). Regarding the synonymous codon usage bias in the CSFV translation initiation region, some codons showing low tRNA abundance in pigs are more frequently located in the translation initiation region than in the open reading frame of CSFV. Although CSFV, similarly to other RNA viruses, has a high mutation rate in nature, the regulatory features of nucleotide and synonymous codon usage of the IRES element, the nucleotide context surrounding the initiation codon and the translation initiation region in CSFV RNA have been 'branded' in the system of translation initiation to accommodate gene expression mediated by the cap-independent translation mechanism.

**Key words:** Cap-independent translation mechanism; tRNA abundance; Classical swine fever virus; Internal ribosome entry site; Synonymous codon usage; Translation initiation region;

## INTRODUCTION

Classical swine fever is a highly contagious viral disease that occurs in swine and results in great economic losses worldwide. Classical swine fever virus (CSFV) has been named as an important disease by the World Organization for Animal Health (OIE). CSFV belongs to the genus *Pestivirus* of the family Flaviviridae and its genome is a positive sense, single-stranded RNA with a large open reading frame (ORF) (Thiel et al., 1991; Jiang et al., 2013). The CSFV ORF is flanked by a 5'-untranslated region (UTR) and a 3'-UTR (Paton et al., 2000). The internal ribosome entry site (IRES) is located in the CSFV 5'-UTR and plays an important role in the translational initiation of the CSFV ORF (Fletcher and Jackson, 2002). The genomes of some RNA viruses are translated through IRES element-mediated translation, which is distinctly different from cap-dependent translation (Fraser and Doudna, 2007; Balvay et al., 2009). These positive sense viral RNAs function similarly to mRNAs, but they contain the 5'-UTR, which exhibits extensive secondary structure features and multiple AUG codons preceding the initiation codon (Firth and Brierley, 2012). The extensive secondary structure plays an important role in IRES activity during the CSFV life cycle, as some CSFV and host cell proteins combine with the proper secondary structure for the IRES (Xiao et al., 2008; Zhu et al., 2010). The function of CSFV IRES is strongly influenced by the coding sequences immediately downstream of the initiation codon (Fletcher and Jackson, 2002). The translation initiation region (TIR), which flanks the initiation codon (AUG), plays an important role in gene expression (Kozak, 2001; Kochetov, 2005). A previous study revealed that the stable RNA structure within the 5'-proximal coding sequence (CDS) does not alter the capacity of the IRES to shape initial contacts with the 40S subunit. However, it may prevent subsequent interactions between the 40S subunit and viral RNA near the initiation codon, which are essential for efficient ribosome entry (Rijnbrand et al., 2001). Some recent studies have reported that synonymous codon usage patterns can control the speed of ribosome scanning along the coding sequence and that synonymous codon usage bias influences translation initiation efficiency at the 5'-termination of the target coding sequence in eukaryotic cells (Cannarozzi et al., 2010; Fredrick and Ibba, 2010; Tuller et al., 2010; Plotkin and Kudla, 2011). Regarding the genetic features of synonymous codon usage in viral coding sequences, some synonymous features

were more frequently selected in a specific region of 5'-proximal CDS and may play a role in regulating the translation of target proteins (Zhou et al., 2013a,c). In this study, we examined the genetic features of the CSFV IRES element, the nucleotides flanking the initiation codon, and the TIR. Our goal was to understand the genetic features of the entire translation initiation system composed of the IRES element, the nucleotides surrounding the initiation codon, and the TIR of CSFV RNA.

## MATERIAL AND METHODS

### Complete genomes of CSFV and tRNA copy numbers in pigs

The complete genomes of 46 CSFV strains were downloaded from the National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov/Genbank/>) (**Table S1**). To investigate the genetic features of IRES and ORFs in the CSFV genome, IRES and ORFs were separated from each genome of CSFV based on the information present at the IRES site of the CSFV genome (AF092448) (Zhu et al., 2010). In addition, to analyze the extent of tRNA pairing to a particular synonymous codon, the copy numbers of cognate tRNAs from swine were obtained from a tRNA abundance database (<http://lowelab.ucsc.edu/>) (**Table S2**).

### Comparative analysis of CSFV IRES sequences

IRES RNA sequences were extracted from the whole genomes of CSFV through alignment using ClustalW with manual editing. Phylogenetic analysis of IRES RNA sequences was performed using the neighbor-joining method with MEGA 4.0 for Windows.

### Identification of 18 consensus nucleotides (n9AUGn)

Eighteen nucleotides, including 9 nucleotides before the initiation codon, 3 nucleotides of the initiation codon, and 6 nucleotides following the initiation codon, of each ORF were chosen via manual editing, where 'n' represents an unknown nucleotide. This procedure was carried out based on a previous report (Crooks et al., 2004).

### Synonymous codon usage in the TIR of CSFV RNA

To analyze synonymous codon usage bias in the TIR composed of aligned codons located in the first 10, 20, 30, 40, and 50 codon positions of the CSFV ORF, we used simple methods based on previous reports (Zhou et al., 2011) using the following equation:

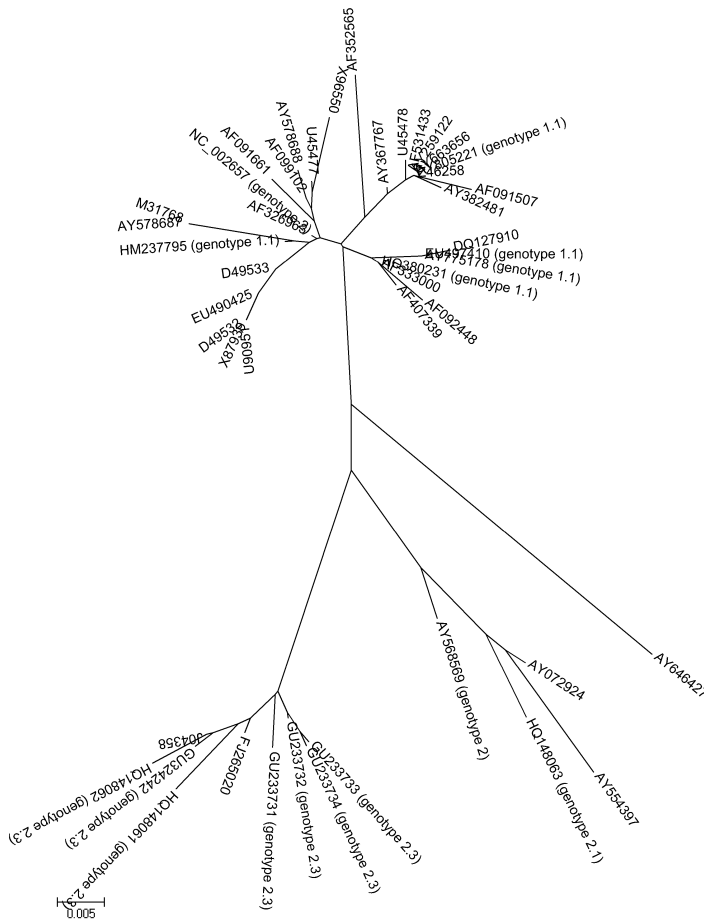
$$Y = \ln\left(\frac{f_n / F_n}{f / F}\right) \quad (\text{Equation 1})$$

where  $f_n$  is the sum of a particular synonymous codon in a specific length ranging from the start codon (AUG) to the  $n$ th codon in the CSFV ORF,  $F_n$  is the sum of the corresponding amino acids in the given region,  $f$  is the sum of this synonymous codon in the CSFV ORF, and  $F$  is the sum of the corresponding amino acid in the given ORF.

## RESULTS

### Comparative sequence analysis of CSFV IRES sequences

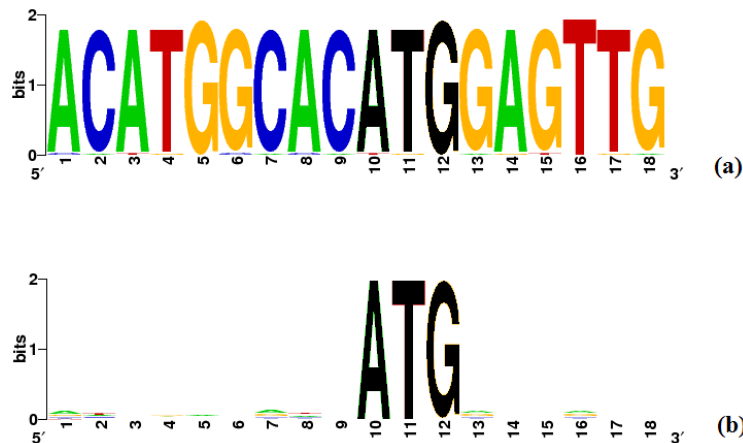
All CSFV IRES element sequences described above were analyzed by preliminary alignment using the ClustalW program. IRES sequences were arranged according to genotype by investigating the associated literature available in GenBank. A phylogenetic clustering tree of CSFV in the IRES RNA sequence was constructed using the neighbor-joining method. Although most IRES sequences were not annotated by genotypes, the genetic diversity of CSFV IRES may represent the separation between genotype 1 and genotype 2. Highly similar sequences of IRES for which genotype assignment was unavailable were generally located in close proximity to IRES sequences of known genotypes (Figure 1).



**Figure 1.** Phylogenetic analysis for the IRES element of CSFV. An unrooted neighbor-joining tree was constructed by the marked genotype of IRES element and the unmarked genotype of IRES element. The bar represents the genetic distance.

### Nucleotide contexts near the CSFV initiation codon

The nucleotide context surrounding the initiation codon or non-initiation codon AUG was compared to understand the significance of nucleotide selection in the region flanking the initiation codon of the CSFV ORF (Figure 2). Compared with the nucleotides surrounding the non-initiation codon AUG in the CSFV ORF, the nucleotides flanking the initiation codon were constrained under translation selection; in contrast, no bias was observed in the nucleotide context surrounding the internal AUG codon (Figure 2). G was found to be the preferred nucleotide (+4 position) following the initiation codon AUG (Figure 2a). The +4G in the nucleotide following the initiation codon of the CSFV ORF is a key position in the Kozak sequence for identifying the translation initiation site (Kozak, 1989). Further analysis showed that the 6 nucleotide positions downstream of the initiation codon of the CSFV ORF have a strong tendency to select GAG UUG, which can encode 2 amino acids: Glu and Leu (Figure 2a). Analysis of tRNA abundance in pigs (Table S2) showed that tRNA copy numbers pairing with GAA were more frequently observed than GAG; however, the synonymous codon GAG (Glu) with a relatively low tRNA abundance was generally selected for the second codon position, closely following the initial codon of the CSFV ORF. The tRNA abundance of UUG, which was frequently selected at the 3rd codon position of the CSFV ORF, is similar to that of GAG (Glu) (Table S2).



**Figure 2.** Comparison of nucleotide context surrounding initiation codon with that of AUG (non-initiation codon) scattering along the CSFV ORF. (a) Nucleotide context surrounding initiation codon. (b) Nucleotide context surrounding non-initiation codon.

### Synonymous codon usage bias in the CSFV TIR

For the 46 CSFV strains examined in this study, synonymous codons were generally selected non-randomly in the TIR. Synonymous codons for the 3 amino acids Phe, Tyr, and His were rarely selected in the first 10 codon positions of the TIR, while the remaining positions of the TIR showed variable tendencies to select these synonymous codons (Figure S1). Synonymous codons for Asn were generally selected with low frequency along the CSFV TIR, and the synonymous codons CAA (Gln) and AAG (Lys) showed a stronger tendency to exist in the TIR than CAG (Gln) and AAG (Lys) (Figure S2). For the synonymous codon usage for

the 3 amino acids Asp, Glu, and Cys, the synonymous codon GAA (Glu) showed higher usage bias than the remaining codons along the CSFV TIR ([Figure S3](#)); however, the synonymous codon GAG (Glu) showed the strongest usage bias in the 2nd codon position, which closely follows the initiation codon in the TIR ([Figure 2a](#)). Based on the tRNA copy numbers of codons encoding Asp and Glu ([Table S2](#)), the synonymous codon GAC (Asp), with relatively high tRNA abundance, was generally observed in the first 10 codon positions; in contrast, the synonymous codon GAA (Glu) showed a weak tendency to be present in the target region ([Figure S3](#)). Val, Ala, and Gly synonymous codons generally existed in the first 10 codon positions in the CSFV TIR, and the synonymous codons GUU and GGC were never selected in the TIR ([Figures S4-S6](#)). For Pro, the synonymous codon CCC did not exist in the CSFV TIR, and the synonymous codon CCA was generally selected in this region ([Figure S7](#)). For Ile, the synonymous codon AUU did not exist in the TIR, and the 3 synonymous members were not selected in the first 20 codon positions ([Figure S8](#)). According to the synonymous codons usage bias for Thr in the CSFV TIR, the 3 members (ACU, ACG, and ACC) did not exist in the first 10 codon positions, and the synonymous codon ACA was generally selected by the TIR ([Figure S9](#)). Based on tRNA abundances pairing to synonymous codons for Arg ([Table S2](#)), the synonymous codon CGG, with the lowest tRNA abundance, was not selected by the TIR, and the synonymous codon CGA represented bias usage in the TIR ([Figure S10](#)). For Leu, the synonymous codon CUC was not observed in the CSFV TIR, while the synonymous members UUA and CUU were found in this region ([Figure S11](#)). Although the synonymous codon UUG (Leu) showed a slight tendency to be present in the first 10 codon positions of the TIR ([Figure S11](#)), this codon showed strong usage bias in the 3rd codon position downstream of the initiation codon ([Figure 2a](#)). For Ser, the synonymous codon AGC showed higher usage bias in the first 10 codon positions of the TIR than the others ([Figure S12](#)). Analysis of the usage of 59 synonymous codons in the CSFV TIR described above and their tRNA abundances ([Figure S1-S12 and Table S2](#)) suggested that the roles of low tRNA abundances are involved in the selection of synonymous codon usage bias in the CSFV TIR. In addition, although the synonymous codons GAG (Glu) and UUG (Leu) showed a weak tendency to be selected for the first 10 codon positions of the TIR, they primarily occupied the 2nd and 3rd codon positions in this region. This suggests that synonymous codon usage in the codon positions in the close downstream region of the initiation codon of CSFV is fixed under viral evolution.

## DISCUSSION

The IRES elements of RNA viruses are generally involved in translation initiation of the viral gene, and their secondary structures play a critical role in this process (Balvay et al., 2009). Phylogenetic analysis of the CSFV 5'-UTR showed that the genetic diversity of the E2, Ns5B, and 5'-UTR genes represent similar clusters (Paton et al., 2000). In this study, although comparison of IRES sequences in CSFV revealed a high level of conservation, a similar phylogenetic cluster of IRES sequences to those of the E2, Ns5B, and 5'-UTR genes was observed. This suggests that recombination between the IRES elements of CSFV strains from different genotypes may be uncommon. CSFV is frequently used as a model system for investigating the genus *Pestivirus*. The 5'-UTRs of this genus represent significant sequence conservation and contribute to the translation and replication of viral RNA genomes (Burks et al., 2011). Although genetic diversity in CSFV IRES elements was observed in this study, high conservation among these IRES sequences results in a stable secondary structure to facilitate



an interaction between IRES elements and viral proteins. NS3 of CSFV can interact with the IRES element to increase translation initiation efficiency (Xiao et al., 2008; Zhu et al., 2010).

The highly conserved nucleotide context surrounding the initiation codon of CSFV may play an important role in translation initiation of the CSFV genome. The 2nd codon position following the initiation codon showed a strong tendency to select the synonymous codon GAG (Glu) pairing in low tRNA abundance, suggesting that this viral genome may prevent inefficient translation initiation by regulating ribosome traffic along its coding sequence. Compared with the specific translation initiation recognition motif of the Kozak sequence (A/GNNATGG), the -3 nucleotide position C in the region flanking the initiation codon of CSFV is at A/G of the Kozak sequence. This suggests that the cap-independent mechanism of translation initiation of CSFV possesses a specific translation initiation region to facilitate translation initiation.

As for the life cycle of viruses, production of viral proteins is often influenced by their host cell environments and translation initiation strategies (Davy and Doorbar, 2007; Jackson et al., 2010; Firth and Brierley, 2012; Zhou et al., 2013b). Synonymous codon usage in the TIR of CSFV has been examined for 59 codons; we found that although the synonymous GAG (Glu) and UUG (Leu) codons showed a slight tendency to exist in the first 10 codon positions of TIR, the 2nd and 3rd codon positions showed a strong tendency to select them. Synonymous codon usage bias for the translation initiation region of a gene is important for regulating the efficiency of translation initiation (Eyre-Walker and Bulmer, 1993; Chen and Inouye, 1994; Stenström et al., 2001a,b). Based on R values, some synonymous codons are highly selected in the TIR, while others are not used as frequently. Codons with clear usage bias in the CSFV TIR may influence the efficiency of translation initiation for this virus. Optimization of the first 5-17 codons of the human chorionic gonadotropin gene increases expression levels by 4- to 5-fold (Vervoort et al., 2000). The usage degree of some synonymous codons in the CSFV TIR was similar to the overall usage degree of the same codons of its host. These codons likely reduce the probability of abortive translation initiation of the CSFV genome through adaptation of codon usage patterns in this region to the host environment. The adaptation of codon usage patterns in exogenous genes to the overall codon usage pattern of the hosts is an effective approach for producing functional proteins (Gustafsson et al., 2004; Burns et al., 2006; Hershberg and Petrov, 2008; Sharp et al., 2010).

## ACKNOWLEDGMENTS

Research supported by grants from the National Natural Science Foundation of China (#31160033), the central universities deriving from the Northwest University for Nationalities (#ZY2012067), and the university's scientific fund for the middle-aged and young scientists deriving from the Northwest University for Nationalities (#12XB37).

## [Supplementary material](#)

## REFERENCES

- Balvay L, Soto Rifo R, Ricci EP, Decimo D, et al. (2009). Structural and functional diversity of viral IRESes. *Biochim. Biophys. Acta* 1789: 542-557.
- Burks JM, Zwieb C, Müller F, Wower IK, et al. (2011). Comparative structural studies of bovine viral diarrhea virus IRES RNA. *Virus Res.* 160: 136-142.
- Burns CC, Shaw J, Campagnoli R, Jorba J, et al. (2006). Modulation of poliovirus replicative fitness in HeLa cells by

- deoptimization of synonymous codon usage in the capsid region. *J. Virol.* 80: 3259-3272.
- Cannarozzi G, Schraudolph NN, Faty M, von Rohr P, et al. (2010). A role for codon order in translation dynamics. *Cell* 141: 355-367.
- Chen GT and Inouye M (1994). Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. *Genes Dev.* 8: 2641-2652.
- Crooks GE, Hon G, Chandonia JM and Brenner SE (2004). WebLogo: a sequence logo generator. *Genome Res.* 14: 1188-1190.
- Davy C and Doorbar J (2007). G2/M cell cycle arrest in the life cycle of viruses. *Virology* 368: 219-226.
- Eyre-Walker A and Bulmer M (1993). Reduced synonymous substitution rate at the start of enterobacterial genes. *Nucleic Acids Res.* 21: 4599-4603.
- Firth AE and Brierley I (2012). Non-canonical translation in RNA viruses. *J. Gen. Virol.* 93: 1385-1409.
- Fletcher SP and Jackson RJ (2002). Pestivirus internal ribosome entry site (IRES) structure and function: elements in the 5' untranslated region important for IRES function. *J. Virol.* 76: 5024-5033.
- Fraser CS and Doudna JA (2007). Structural and mechanistic insights into hepatitis C viral translation initiation. *Nat. Rev. Microbiol.* 5: 29-38.
- Fredrick K and Ibbas M (2010). How the sequence of a gene can tune its translation. *Cell* 141: 227-229.
- Gustafsson C, Govindarajan S and Minshull J (2004). Codon bias and heterologous protein expression. *Trends Biotechnol.* 22: 346-353.
- Hershberg R and Petrov DA (2008). Selection on codon bias. *Annu. Rev. Genet.* 42: 287-299.
- Jackson RJ, Hellen CU and Pestova TV (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11: 113-127.
- Jiang DL, Gong WJ, Li RC, Liu GH, et al. (2013). Phylogenetic analysis using E2 gene of classical swine fever virus reveals a new subgenotype in China. *Infect. Genet. Evol.* 17: 231-238.
- Kochetov AV (2005). AUG codons at the beginning of protein coding sequences are frequent in eukaryotic mRNAs with a suboptimal start codon context. *Bioinformatics* 21: 837-840.
- Kozak M (1989). The scanning model for translation: an update. *J. Cell Biol.* 108: 229-241.
- Kozak M (2001). New ways of initiating translation in eukaryotes? *Mol. Cell. Biol.* 21: 1899-1907.
- Paton DJ, McGoldrick A, Greiser-Wilke I, Parchariyanon S, et al. (2000). Genetic typing of classical swine fever virus. *Vet. Microbiol.* 73: 137-157.
- Plotkin JB and Kudla G (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* 12: 32-42.
- Rijnbrand R, Bredenbeek PJ, Haasnoot PC, Kieft JS, et al. (2001). The influence of downstream protein-coding sequence on internal ribosome entry on hepatitis C virus and other flavivirus RNAs. *RNA* 7: 585-597.
- Sharp PM, Emery LR and Zeng K (2010). Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 365: 1203-1212.
- Stenström CM, Holmgren E and Isaksson LA (2001a). Cooperative effects by the initiation codon and its flanking regions on translation initiation. *Gene* 273: 259-265.
- Stenström CM, Jin H, Major LL, Tate WP, et al. (2001b). Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in *Escherichia coli*. *Gene* 263: 273-284.
- Thiel HJ, Stark R, Weiland E, Rumenapf T, et al. (1991). Hog cholera virus: molecular composition of virions from a pestivirus. *J. Virol.* 65: 4705-4712.
- Tuller T, Carmi A, Vestsigian K, Navon S, et al. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141: 344-354.
- Vervoort EB, van Ravestein A, van Peij NN, Heikoop JC, et al. (2000). Optimizing heterologous expression in dictyostelium: importance of 5' codon adaptation. *Nucleic Acids Res.* 28: 2069-2074.
- Xiao M, Bai Y, Xu H, Geng X, et al. (2008). Effect of NS3 and NS5B proteins on classical swine fever virus internal ribosome entry site-mediated translation and its host cellular translation. *J. Gen. Virol.* 89: 994-999.
- Zhou JH, Zhang J, Chen HT, Ma LN, et al. (2011). The codon usage model of the context flanking each cleavage site in the polyprotein of foot-and-mouth disease virus. *Infect. Genet. Evol.* 11: 1815-1819.
- Zhou JH, Gao ZL, Sun DJ, Ding YZ, et al. (2013a). A comparative analysis on the synonymous codon usage pattern in viral functional genes and their translational initiation region of ASFV. *Virus Genes* 46: 271-279.
- Zhou JH, Su JH, Chen HT, Zhang J, et al. (2013b). Clustering of low usage codons in the translation initiation region of hepatitis C virus. *Infect. Genet. Evol.* 18: 8-12.
- Zhou JH, Zhang J, Sun DJ, Ma Q, et al. (2013c). Potential roles of synonymous codon usage and tRNA concentration in hosts on the two initiation regions of foot-and-mouth disease virus RNA. *Virus Res.* 176: 298-302.
- Zhu Z, Wang Y, Yu J, Wan L, et al. (2010). Classical swine fever virus NS3 is an IRES-binding protein and increases IRES-dependent translation. *Virus Res.* 153: 106-112.