Infection, Genetics and Evolution 34 (2015) 17–25

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Analysis of codon usage pattern evolution in avian rotaviruses and their preferred host

Jobin Jose Kattoor[a], Yashpal Singh Malik[a,*], Aravind Sasidharan[b], Vishnuraj Mangalathu Rajan[a], Kuldeep Dhama[a], Souvik Ghosh[c], Krisztián Bányai[d], Nobumichi Kobayashi[e], Raj Kumar Singh[a]

[a] Indian Veterinary Research Institute, Izatnagar 243 122, Uttar Pradesh, India
[b] National Institute of High Security Animal Diseases, Bhopal, Madhya Pradesh, India
[c] Department of Biomedical Sciences, One Health Center for Zoonoses and Tropical Veterinary Medicine, Ross University School of Veterinary Medicine, P. O. Box 334, Basseterre, Saint Kitts and Nevis
[d] Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Hungária krt. 21, Budapest 1143, Hungary
[e] Sapporo Medical University School of Medicine, Chuo-Ku, Sapporo, Japan

ARTICLE INFO

Article history:
Received 28 February 2015
Received in revised form 26 May 2015
Accepted 14 June 2015
Available online 16 June 2015

Keywords:
Avian rotavirus
VP6 gene
Relative synonymous codon usage
Effective number of codons
Codon adaptation index
Geographical separation
Virus evolution in host

ABSTRACT

Rotavirus infection is a worldwide problem, with occurrence of highly divergent viruses classified in 8 species (A–H). We report here the evolution assessment of codon usage patterns in virus–host system in avian rotavirus (AvRV) of species RVA, RVD, RVF and RVG (preferentially affecting birds). The nucleotide contents, codon usage bias (CUB), relative synonymous codon usage (RSCU), and effective number of codons (ENCs) values were investigated targeting overexpressing major inner capsid viral protein (VP6) of these AvRV species. The results confirm that the evolutionary characteristics influences the rotavirus (RV) genetic diversity and impact of host's natural selection on the AvRVs codons. Synonymous codon usage patterns were evaluated following multivariate statistical procedures on all available AvRV coding gene sequences. RSCU trees accommodated all AvRV species and preferred host sequences in one topology confirming greater imminence of AvRVs with the host chicken cell genes. Similarly, the codon adaptation index (CAI) results also displayed a higher adaptation of AvRVs to its chicken host. The codon preference analysis of RVs revealed that VP6 gene express more proficiently in the yeast system, whereas, codon optimization might be required for the effectual expression in *Escherichia coli* and *Homo sapiens*. The findings provide basic evidence on the dynamics of AvRV evolution and its host adaptation, which could be exploited for additional research on avian species in future.

1. Introduction

The Codon Usage Bias (CUB) is an exclusive uniqueness of genomic sequences encoding proteins (Hershberg and Petrov, 2008). Hitherto studies proved that certain genes display defined preference for certain codons for the same amino acid, rather using the available codons randomly. The changes at 3rd codon's GC%s and synonymous codons in coding DNA, are the most apparent criterion in codon usage evolution which possibly lead to noticeable CUB (Shackelton et al., 2006; Yang and Nielsen, 2008), making them a mark for studying virus–host evolution. Furthermore, the compositional limits, translational selection (Ghosh et al., 2000) and mutational forces (Jenkins and Holmes, 2003) are other identified factors observed making significant part in the relative synonymous codon usage pattern. The redundancy of genetic code and synonymous codon usage offers the organism an opportunity to focus on efficient protein synthesis and folding (Ikemura, 1982) and possessing the analogous amino acid sequence (Stoletzki and Eyre-Walker, 2007). It has been found that, in unicellular organisms like *Escherichia coli* and *Bacillus subtilis*, highly expressed genes have high preferential usage of certain codons with matching tRNA (Sharp and Li, 1986), although lesser

---

Abbreviations: VP, viral protein; AvRV, avian rotavirus; RVA, rotavirus species A; RVD, rotavirus species D; RVF, rotavirus species F; RVG, rotavirus species G; ORF, open reading frame; CAI, codon adaptation index; CoA, correspondence analysis; ENc, effective number of codons; RSCU, relative synonymous codon usage values; GC_3, the frequencies of nucleotide G + C at third codon position.

* Corresponding author.

E-mail addresses: jobinjkattoor@gmail.com (J.J. Kattoor), malikys@ivri.res.in (Y.S. Malik), aravindspillai02@gmail.com (A. Sasidharan), vishnurajmr@gmail.com (V.M. Rajan), dhamak@rediffmail.com (K. Dhama), souvikrota@gmail.com (S. Ghosh), bkrota@hotmail.com (K. Bányai), koba103161@yahoo.co.jp (N. Kobayashi), rks_virology@rediffmail.com (R.K. Singh).

http://dx.doi.org/10.1016/j.meegid.2015.06.018
1567-1348/© 2015 Elsevier B.V. All rights reserved.

expressing genes have virtually identical synonymous codon usage for a specific amino acid (Lesnik et al., 2000).

Rotaviruses (RVs) are the most common causes of sporadic or epidemic gastroenteritis among several mammalian and avian species (Mata et al., 1983; Shukry, 1986; Dhama et al., 2009; Siqueira et al., 2010; Kattoor et al., 2013; Malik et al., 2014). The virus belongs to the family *Reoviridae* and virions are non-enveloped consisting of 11 double stranded RNA genome segments (Estes and Kapikian, 2007). Most segments are having one open reading frame (ORF), encoding for a single protein, except the 11th genome segment which encodes for two polypeptides (NSP5/6). The virus structure is triple-layered and possess icosahedral symmetry. The outer layer is made up of VP4 and VP7 proteins, while VP6 protein composes intermediate layer. The serological and genomic analysis classifies RV into eight species (RVA to RVH) (Estes and Kapikian, 2007; Molinari et al., 2014). Hitherto analysis revealed availability of ample information on human, porcine and bovine RVs, however, information on RVs of avian species is scanty. Though reports confirm that RV species D, F and G selectively infect and causes economic loss in poultry industry, studies are in particular limited elucidating the host preferences for the AvRVs (Otto et al., 2012). The genome of AvRVs is relatively small and these viruses mainly depend on their host for replication, necessitates the understanding of codon usage biasness and the evolution of these viruses to know adaptation for a particular host. We report here a comprehensive analysis of CUB for different species of AvRVs (RVA, RVD, RVF and RVG) which are known to affect domestic poultry. In this study, we evaluated the subtleties of virus evolution and adaptation to its host by reviewing the nucleotide contents and effective number of codons (ENCs), relative synonymous codon usage (RSCU), and codon usage bias (CUB) values.

2. Materials and methods

2.1. Sequence data retrieval

Accessible VP6 gene sequences of avian RVD ($n = 15$) and representative VP6 gene sequences of avian RVA ($n = 3$), RVF ($n = 1$) and RVG ($n = 1$) were recuperated from the National Center for Biotechnology Information (NCBI) GenBank database (http://www.ncbi.nlm.nih.gov/Genbank/) for analysis of various parameters of codon usage biasness. The detailed information about the viruses including accession numbers and strain names are listed in Supplementary Table 1. Since the number of complete genomes for AvRV available in GenBank did not yield factual diversity, gene sequences with partial length were also included in addition to the complete genomes. The positions indicated are according to the VP6 gene sequence of the one of our avian RVD strains, named UKD 48 (GenBank Accession No: JX187435) isolated from India during 2012.

2.2. Codon usage bias analysis

The parameters studied for codon usage bias analysis includes (i) Compositional properties measures specifically for general nucleotide composition as well as nucleotide composition at 3rd position of codon, (ii) relative synonymous codon usage (RSCU), (iii) effective number of codons (ENc) usage and (iv) codon adaptation index (CAI). The Biosoftware DNAStar 7.0 and online Mobyle Pasteur software interface (http://mobyle.pasteur.fr/) were used to assess the CUB parameters of AvRV VP6 coding sequences. RSCU value was calculated as described by Roth et al. (2012). SPSS 13.0 software was used for the analysis of codon usage bias in AvRV VP6 gene. It was equated with *E. coli*, yeast and *Homo sapiens* using the codon usage information available to assess the preferential codon usage for *E. coli*, yeast and *H. sapiens* (http://www.kazusa.or.jp/codon).

2.3. Statistical analysis

The Pearson's rank method was implemented to apprehend the codon usage pattern of AvRVs and relationship between general nucleotide composition and nucleotide composition at 3rd codon position in the VP6 gene coding sequence. Spearman's rank correlation analysis was also done. Statistical analysis of the results was carried out by using SPSS software (version 13; SPSS, Inc., Chicago, IL).

3. Results

3.1. Synonymous codon usage pattern in AvRV VP6 gene

The compositional analysis of the VP6 gene of AvRVs total nucleotide contents revealed that the A% (32.59) and U% (29.16) were distinctly greater than G% (18.35) and C% (19.89). The results of nucleotide composition analysis at 3rd position reveled that A_3% (44.27) and U_3% (36.62) were higher than G_3% (24.35) and C_3% (23.62). To measure the magnitude of codon bias in AvRV, the average values for all the triplets and the codon preferences of VP6 gene were considered during the study. The GC content varied between 38.02% and 41.5% (mean-39.5%; standard deviation-0.01), specifying that nucleotide G and C are the inconsequential elements of the AvRV VP6 gene.

The results also confirmed that the RSCU values of most of the highly preferred codons used for amino acids ended with A/U and a large percentage of less abundant codons ended with G/C. A total number of 7 codons used by different amino acids like Cys (TGC), Glu (GAG), Leu (CTG, TTG), Arg (CGC), Val (GTG) and Tyr (TAC) used codons ending with G/C. It was noted that 6 codons (GCC, CCC, CGG, GCG, GTC and CCT) were not used for the protein product in AvRV VP6 gene. The codon count and the average RSCU values of high and less expressed VP6 gene of different isolates are tabulated in Table 2 and Supplementary Table 3. Seventeen codons (each for an amino acid), including GCA (Ala), TGC (Cys), GAT (Glu), TTT (Phe), GGA (Gly), CAT (His), ATA (Iln), AAA (Lys), CTA (Leu), AAT (Asn), TAC (Tyr), GTA (Val), ACA (Thr), TCA (Ser), AGA (Arg), CAA (Gln) and CCA (Pro) were observed as the ideal codons. Specially, in case of Glu, GAG codon showed higher RSCU value but GAA scored highest frequency. Tyr also showing the same pattern TAC has highest RCSU value but shows less frequency and TAT has highest frequency. This further strengthens the results of A/T ending of preferred codons.

Moreover, a relative study of the RSCU values among AvRV and chicken cell (Rao et al., 2011) was also performed (Supplementary Table 3), where it was supported that the codon usage pattern of AvRV was typically agreeing with that of ideal host. We found certain under-represented codons as well as over-represented codons. Ala, Glu, Asp, Gly, Leu, Try, Val, Thr, Arg and Pro showed the similar synonymous codon usage pattern. This similarity in the codon usage pattern between host and virus could support in the translational efficiency of the analogous amino acids, by adjusting to its host beneath translation selection. Few of the uncommon codons in chicken cells like TTA for Leu, TCG for Ser and CCG for Pro were seen in RV genome.

3.2. Influencing elements of codon usage pattern

The correlation analysis results obtained using Pearson's rank correlation assessment showed significant positive correlation between GC_3s (0.627), GC contents (0.907), hydrophobicity