Infection, Genetics and Evolution 32 (2015) 425–431

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Short communication

Genome analysis of a novel, highly divergent picornavirus from common kestrel (*Falco tinnunculus*): The first non-enteroviral picornavirus with type-I-like IRES

Ákos Boros [a], Péter Pankovics [a], Peter Simmonds [b], Edit Pollák [c], Róbert Mátics [d,e], Tung Gia Phan [f,g], Eric Delwart [f,g], Gábor Reuter [a,*]

[a] Regional Laboratory of Virology, National Reference Laboratory of Gastroenteric Viruses, ÁNTSZ Regional Institute of State Public Health Service, Pécs, Hungary
[b] Infection and Immunity Division, Roslin Institute, University of Edinburgh, Edinburgh, UK
[c] Department of General Zoology, Faculty of Sciences, University of Pécs, Hungary
[d] Department of Pathophysiology, University of Pécs Medical Center, Hungary
[e] Hungarian Nature Research Society (HuNaReS), Ajka, Hungary
[f] Blood Systems Research Institute, San Francisco, CA, USA
[g] University of California, San Francisco, CA, USA

ARTICLE INFO

Article history:
Received 11 December 2014
Received in revised form 2 April 2015
Accepted 3 April 2015
Available online 9 April 2015

Keywords:
Kestrel
Falcon
Picornavirus
IRES
Enterovirus
Avian

ABSTRACT

Although the number of identified avian-borne picornaviruses (family *Picornaviridae*) is continuously increasing there remains several species-rich avian host groups, such as the order Falconiformes (with 290 bird species) from which picornaviruses have not been identified. This study reports the first complete genome of a novel, highly divergent picornavirus, named as Falcovirus A1 (KP230449), from the carnivorous bird, the common kestrel (*Falco tinnunculus*, order Falconiformes). Falcovirus A1 has the longest $3D^{RdRp}$ genome region and distant phylogenetic relationship to the Hepatitis A virus 1 (*Hepatovirus*) and Avian encephalomyelitis virus 1 (*Tremovirus*). It has a type-I (enterovirus-like) IRES in the 5′UTR – identified for the first time among avian-borne picornaviruses suggesting that type-I IRES is not restricted only to enteroviruses and providing further evidence of mosaicism of this region among different picornavirus genera.

1. Introduction

Members of the family *Picornaviridae* are small viruses with positive sense, single-stranded RNA genome. The picornaviral genomes, ranging in size from 7.2 to 9.7 kilobases, generally follow a characteristic organizational pattern: VPg-5′UTR-(L)-P1(VP4-2-3-1-P2(2ABC)-P3(3ABCD)-3′UTR-poly(A)-tail (Palmenberg et al., 2010). Picornaviruses differ, however, in the presence or absence of an N-terminal Leader protein which could belong to multiple types, in the nature of the VP4/VP2 and VP1/2A cleavage sites, the function of 2A and in the number of encoded 2A and 3B proteins (Gorbalenya and Lauber, 2010; Knowles et al., 2012; Boros et al., 2014).

The family *Picornaviridae* currently consists of 50 species grouped into 29 genera (Knowles et al., 2012; Adams et al., 2013; www.picornaviridae.com).

According to the phylogenetic distribution in the most conservative 3D RNA-dependent RNA polymerase (RdRp) region, several different main picornavirus clades/lineages (sequence of monophyletic genera clustered together), i.e. enterovirus, cardiovirus, kobuvirus, parechovirus and hepatovirus can be identifiable, and all of them contain picornaviruses identified from avian source (“avian-borne picornaviruses”) (Lau et al., 2014; Boros et al., 2014).

All picornavirus genomes contain a highly structured, secondary RNA structure at the 5′UTR genome region containing an internal ribosomal entry site (IRES). Structurally and functionally, these can be divided into 5 types (I–V; Palmenberg et al., 2010; Sweeney et al., 2012) that show varying distributions among different picornavirus genera in a patterns that indicates multiple exchanges of this element between virus groups in their evolutionary history (mosaicism). For example, the type-II IRES is found

* Corresponding author at: Regional Laboratory of Virology, National Reference Laboratory of Gastroenteric Viruses, ÁNTSZ Regional Institute of State Public Health Service, Szabadság út 7, H-7623 Pécs, Hungary. Tel.: +36 06 (72) 514 979.
E-mail address: reuter.gabor@ddr.antsz.hu (G. Reuter).

http://dx.doi.org/10.1016/j.meegid.2015.04.005
1567-1348/

among the members of the *Parechovirus*, *Rosavirus* and *Hunnivirus* genera which are forming separate phylogenetic groups, while the type-IV IRES is found in sapeloviruses, aquamaviruses, duck hepatitis A viruses and teschoviruses, as well as some members of a different virus family (*Flaviviridae*) (Reuter et al., 2012; Phan et al., 2013b; Hellen and de Breyne, 2007). However, to our current knowledge, the type-I (*Enterovirus*), type-III (*Hepatovirus*) and type-V (*Aichivirus A* and *B*/*Passerivirus*/*Salivirus*) IRESes are entirely or almost entirely specific to certain lineages and distributed only among phylogenetically related viruses (Palmenberg et al., 2010; Sweeney et al., 2012; Lau et al., 2011).

Although to date, most picornaviruses have been identified in mammalian hosts; their wide distribution among different orders of birds is increasingly recognised. To date the identified avian picornaviruses are belong to at least 17 species grouped into 13 genera which currently form five phylogenetic clusters, from which the tremovirus cluster was the only one which contains only one member, the Avian encephalomyeleitis virus 1 (AEV-1) of genus *Tremovirus*. Majority of the known avian picornaviruses were identified from poultry birds (quail, turkey, chicken, and ducks) (Boros et al., 2014). To our currently knowledge, no picornaviruses have been detected in carnivorous birds and the common kestrels (*Falco tinnunculus*) of the family Falconidae has not been investigated before. This study reports the complete sequence and analysis of a novel, highly divergent picornavirus identified from common kestrel, with potential type-I IRES and unusually long $3D^{RdRp}$ genome region. This virus was named as Falcovirus A1 (FaV-A1) which could be the founding member of a novel genus "Falcovirus" (the name refers to the scientific name of common kestrel (*F. tinnunculus*).

2. Materials and methods

A single cloacal sample (VOVE0622) was collected during the regular bird ringing process from an apparently asymptomatic, approximately 5-week-old common kestrel (*F. tinnunculus*) from a small nest near village Harkakötöny (Bács-Kiskun County, Southeast Hungary) in June 2013.

The viral metagenomic analysis of the VOVE0622 sample was conducted using sequence independent random RT-PCR amplification of viral-particle protected nucleic acids. For the production of viral cDNA library ScriptSeqTM v2 RNA-Seq Library Preparation Kit (Epicentre) was used. The library was sequenced by the Miseq Illumina platform, as described previously (Phan et al., 2013a). For the determination of the complete picornavirus genome (kestrel/VOVE0622/2013/HUN; KP230449) and for the verification of the metagenomic contigs long-range and conventional RT-PCR amplification, 5′/3′ rapid amplification of cDNA ends (RACE) and dye-terminator sequencing were used as previously described (Boros et al., 2011, 2012). The contigs from the metagenomic experiment served as templates for the sequence-specific primer design. The primers used for the RT-PCR based amplification and sequencing were available on request. Generic $3D^{RdRp}$ primers (KestrelPV-3D-F: 5′ TGT ATC ATA TCT CAA GGA TGA A 3′ and KestrelPV-3D-R: 5′ AAG ACA AGA TCA TCC CCA TA 3′) are designed targeting the $3D^{RdRp}$ genome-region of kestrel/VOVE0622/2013/HUN for screening additional cloacal samples ($N = 14$) collected from apparently healthy common kestrels in Hungary in 2013. Samples including VOVE0622 were collected by qualified ornithologists with valid permission (Permit No. of the National Inspectorate for Environment, Nature and Water: 14/131-7/2012 and 14/3858-9/2012).

The amino acid (aa) pairwise alignments and identity calculations were performed by the BioEdit software (version 7.1.3.0) (Hall, 1999) using the in-built ClustalW algorithm (Thompson et al., 1994). The aa alignments for the phylogenetic reconstruction and cleavage site analysis were generated using the MUSCLE algorithm (Edgar, 2004). The possible proteolytic cleavage sites were predicted by the NetPicoRNA web tool, where the polyprotein of the study sequence was used as template (Blom et al., 1996). The aa phylogenetic trees were constructed using the Maximum likelihood method with two substitution models: Le_Gascule_2008 model (LG) with Freqs and gamma distributed, invariant sites (G + I) in case of P1 and LG with G + I in case of 3D of MEGA software (ver. 6.06) (Tamura et al., 2013). The substitution models were selected based on the results of the Best Model search of MEGA 6.06. Bootstrap (BS) values (based on 1000 replicates) for each node are shown if BS > 50%. The list of viruses and accession numbers used for the phylogenetic analysis is provided in Supplementary Material Table S3. The secondary RNA structure of the 5′ and 3′UTR regions were predicted by the Mfold program (Zucker, 2003), visualized using the Corel Draw Graphic Suite (Ver. 12).

A list of the accession numbers used for the nucleotide composition analysis (NCA) is provided in Supplementary Material Table S1. Mononucleotide and dinucleotide frequencies for each sequence were determined using the program "Composition Scan" in the SSE version 2.1 (Simmonds, 2012). Dinucleotide biases were determined as the ratio between the observed frequencies of each of the 16 dinucleotides from the expected frequencies determined by multiplying the frequencies of each of the two constituent mononucleotides. Discriminant analysis was performed using the statistical package, SYSTAT with default parameters as previously described (Kapoor et al., 2010). Sequences from positive stranded RNA viruses were assigned to three host categories, vertebrate, invertebrate and plant and frequencies of each mononucleotide and dinucleotide used as predictive factors to infer host ranges of unknown virus sequences from the current study.

3. Results

The *in silico* analysis of the sequence reads of the viral metagenomic analysis resulted a 143- and a 201-nt-long non-overlapping sequences (Fig. 1) showed closest sequence similarity to the VP3 genome region of Cosavirus A (35% aa identity to ADJ39692) and to the 2C genome region of hepatitis A virus (50% aa identity to BAC16580) as the closest matches using GenBank BLASTx search. Based on these two picornavirus sequences the complete genome of a Falcovirus A1 (FaV-A1) strain kestrel/VOVE0622/2013/HUN, (GenBank ID: KP230449) was characterised using RT-PCR methods. The continuous 8003-nucleotide(nt)-long RNA genome of the FaV-A1 was predicted to have a picornavirus-like genome organization pattern: 5′UTR-P1(VP4-VP2-VP3-VP1)-P2(2A-2B-2C)-P3(3A-3B-3C-3D)-3′UTR (Fig. 1A).

The genome regions of P1 (2583 nt; 861 aa), P2 (1851 nt; 617 aa) and P3 (2829 nt; 942 aa) showed substantial divergence from the representative members of the official and unassigned picornavirus genera (Table 1). The potential proteolytic cleavage sites of FaV-A1 mapped based on the results of NetPicoRNA search (Table 2) with the exception of VP4/VP2 and 2C/3A sites, where the alignment with the representative member of the closest relative (Hepatitis A virus) was used for the predictions (data not shown). The closest relative was selected based on the results of BLASTx search (Table 2), where the VP0 and 2C genome regions of FaV-A1 were used as query sequences. Based on the closest similarity of the study sequence with the hepatitis A virus variant CY145 (AAA45473) in the VP0 (VP4 and VP2) region (Table 2) an unusual, non-$3C^{protease}$ cleavage site of VP4/VP2 (DAARLA↓QR) was predictable similar as found in hepatitis A viruses (HILSLA↓DI, identical aas were underlined) (Cohen et al., 1987). Due to the considerable sequence divergence between FaV-A1 and the other members of the *Picornaviridae* all the mapped