

**Identification and characterization of novel picornaviruses
by molecular biological methods**

Ph.D. Thesis

Péter Pankovics

Supervisor: ¹Gábor Reuter, M.D., Ph.D., Med. Habil.

Program leader: ²Levente Emödy, M.D., D.Sc.

¹Government Office for Baranya County,

Regional Laboratory of Virology,

National Reference Laboratory of Gastroenteric Viruses, Pécs

²Department of Medical Microbiology and Immunology, Faculty of Medicine

University of Pécs



Pécs

2013.

TABLE OF CONTENTS

| | |
|---|----|
| INTRODUCTION | 2 |
| AIMS | 5 |
| MATERIALS AND METHODS | 6 |
| RESULTS | 7 |
| DISCUSSION AND NOVEL FINDINGS | 12 |
| PUBLICATIONS SERVED AS BASIS FOR THE THESIS | 15 |

INTRODUCTION

The picornaviruses (family *Picornaviridae*) are highly wide-spread in living world causing numerous animal and human diseases with diversified morbidity and mortality. To our current knowledge this virus family seems to contain the most numerous virus species, being present from fish to primates. The National Committee in Taxonomy of Viruses (ICTV) has separated 12 genera since 2008 in the family: *Aphthovirus*, *Avihepatovirus*, *Cardiovirus*, *Enterovirus*, *Erbovirus*, *Hepatovirus*, *Kobuvirus*, *Parechovirus*, *Sapelovirus*, *Senecavirus*, *Teschovirus* and *Tremovirus*. The family has extended to 5 more genera since 2013: *Aquamavirus*, *Cosavirus*, *Dicipivirus*, *Megrivirus* and *Salivirus*. Each and every genus consists of numerous virus species.

The picornaviruses are small, 22 – 30 nm in diameter non-enveloped viruses with positive, single-stranded RNA (+ssRNA) genome. The viral genome has 7000-9700 nucleotides (nt) in length, having one open reading frame (ORF), from which classically one polyprotein is translated. Untranslated regions (UTRs) are found at both ends of the genome. The 5'UTR of the picornaviruses is various in lengths and can be described with complex secondary RNA structure. Within this region, from the point of view of function and structure, the internal ribosome binding site (IRES) should be highlighted, which currently has 5 types (IRES I-V). The 3'end of the picornavirus genome with a secondary-RNA-structured untranslated region (3'UTR) in various lengths ending in a polyadenine tail can also be found.

The coding regions of picornaviruses are divided into 3 permanent regions: P1, P2, and P3. Upstream the P1 region some picornaviruses are coding a so-called Leader protein (L-protein). The P1 region encodes the capsid proteins, emphasizing the VP1, which is responsible for the antigenicity. The protein products of P2 and P3 regions play an indispensable role in protein cleavages (2A^{pro}, 3C^{pro}, 3CD^{pro}) and in the replication of viral genome (2B, 2C, 3AB, 3B^{VPg}, 3CD^{pro}, 3D^{pol}).

The picornaviruses, which are known to cause human diseases (*Enterovirus*, *Hepatovirus*, *Kobuvirus* and *Parechovirus*) can cause different clinical syndromes (common cold, poliomyelitis, meningitis, acute flaccid paralysis, hepatitis, gastroenteritis, myocarditis, herpangina, etc.). The transmission of the viruses from one host to another takes place horizontally, where the most important ways are the faecal-oral and the respiratory route.

According to the taxonomical classification of the ICTV *Picornaviridae* Study Group, we can talk about picornavirus members of a particular genus when (1) the viral proteins of L, 2A, 2B and 3A are homologous within the members of the genus, (2) the members of the genus have structurally homologous IRES (identical IRES type) and (3) the P1, P2 and P3 regions of virus species in the same genus must be phylogenetically related to one another and the amino acid (aa) identity of P1, P2 and P3 is more than 40%, 40% and 50%, respectively.

Genus Kobuvirus

A novel viral pathogen from stool samples collected from a gastroenteritis outbreak associated with shellfish consumption in Aichi Prefecture, Japan was described in 1991. The complete genome of the novel virus species (Aichi virus; since March 2013 called *Aichivirus A*) was successfully determined in 1998, being the first representative prototype in genus *Kobuvirus*. In 2003, the relative of Aichi virus, the bovine kobuvirus (U-1, since March 2013 called *Aichivirus B*) was discovered faecal samples of cattle. The name of kobuvirus came from the Japanese word 'kobu', meaning knob, node, which refers to the visible knobs or

nodes of the surface structures of the virus on the electron micrograph. The kobuviruses possess L-protein.

Based on Japanese, German, French and Spanish seroepidemiological studies, the 80-90% of the population have antibody against Aichi virus by ages of 30-40. The infection of Aichi virus will probably not give long-time immunity; re-infections can be frequent. The kobuviruses are likely to spread by faecal-oral and direct contact routes, by contaminated food or water causing even epidemics.

The Aichi virus is believed to cause disease in humans with diarrhoea. The main clinical symptoms of the infections are stomach and intestinal disorders, diarrhoea, but sometimes abdominal pain, nausea, vomiting and fever also occur.

Before the beginning of our research only a few studies have dealt with the detection of Aichi virus and bovine kobuvirus with molecular biological methods. The Aichi virus was detected rarely (<3%) in sporadic and epidemic gastroenteric diseases. The bovine kobuvirus was only detected in Japan and Thailand and the detection rate was less than 17% in both studies. The virus may be present in endemic form in cattle herds.

Genus Enterovirus

The members of the genus *Enterovirus* (EV) are variable; on one hand there are various types of viruses belonging to this genus (rhinoviruses, polioviruses, enteroviruses, etc.), and on the other, they cause different diseases ranging from asymptomatic or mild upper respiratory tract (common cold), gastroenteral infections to rare but serious illnesses, such as encephalitis/meningitis or acute flaccid paralysis. Based on the most important viral capsid protein, VP1, human enteroviruses (except rhinoviruses) are classified into four viral species: EV-A, EV-B, EV-C and EV-D. The number of newly discovered enteroviruses causing human diseases is rapidly increasing. A novel, recombinant enterovirus (EV-C109), which is classified into species EV-C was discovered in February 2008, in childhood respiratory infections from Nicaragua. The 5'UTR region of

EV-C109 has EV-A origin, while the coding and the 3'UTR regions are originated from EV-C.

Genus Sapelovirus

The genus *Sapelovirus* within *Picornaviridae* has officially been accepted since 2006 as a novel genus. Based on the comparative evolutionary analyses, the genus includes three viral species: simian, avian and porcine sapeloviruses.

AIMS

- Our aims were to identify and characterize the serendipitously discovered (porcine) kobuvirus from faecal samples of domestic swine by molecular genetic methods.
- Our aims were to determine, characterize and phylogenetically analyse the complete nucleotide sequence of porcine kobuvirus. Our aims were to define the secondary structures of the 5'UTR and IRES.
- Our aims were the molecular epidemiological investigation and follow-up of the porcine kobuvirus from swine faecal/serum sample pairs; investigation of the possibility of viraemia of porcine kobuvirus together with the evolutionary analysis based upon two complete viral genomes.
- Our aims were the detection of kobuvirus from faecal samples of other livestock animal species (bovine, ovine, etc.).
- Our aims were the detection of Aichi virus from stool and respiratory samples of children.
- Our aims were to detect kobuvirus in human stool and quail faecal samples using the screening primer pairs, the so-called UNIV-kobu-R/UNIV-kobu-F, which were newly designed to the conservative 3D region of kobuviruses.
- Our aims were to set up and improve the molecular biological/genetic methods (RT-PCR, Long-Range RT-PCR, primer walking PCR, 5'/3' RACE RT-PCR) in our laboratory in order to determine the complete genomes of picornaviruses.

MATERIALS AND METHODS

Specimens

Specimens were collected between 2000 and 2011. Faecal, serum and egg surface samples were collected from animals. Stool samples and nasopharyngeal aspirates were collected from humans. Faecal samples were collected from patients with diarrhoea and from asymptomatic and symptomatic animals. The animal faecal samples derived from domestic pigs (*Sus scrofa domestica*), cattle (*Bos taurus*), ovine (*Ovis aries*) and domestic common quail (*Coturnix coturnix*).

RNA extraction, RT-PCR and 5'/3' RACE methods

Up to 10-50% PBS diluted faecal samples were used. Serum samples and nasopharyngeal aspirates were not diluted. Total RNA was extracted with TRIzol[®]/TRIzol LS[®] reagents (*Serva; Gibco BRL*) according the manufacturer's protocol, then the RNA was stored in -80°C until used. The amplification of the viral RNA was carried out by RT-PCR. Primer walking technique was used for the determination of complete viral genomes; rapid amplification of cDNA ends (5'/3' RACE, *Roche*) method was applied in order to define the 5'/3' regions according to the manufacturer's instruction. The nucleotide sequences of UNIV-kobu-R (5'ATGTTGTTTRATGATGGTGTGA3', antisense, R: A/G) and UNIV-kobu-F (5'TGGAYTACAAGTGTGTTTGGATGC3', sense, Y: C/T), which were the key primer pairs of the investigations, were designed to the conservative 3D^{pol} gene of kobuviruses. The screening and sequence specific primers for the investigations were synthesized by IDT (*Belgium*).

Gel electrophoresis, sequencing and phylogenetic analysis

PCR products were run in 1.5% agarose gel buffered with Tris-Borate-EDTA and stained with ethidium-bromide. PCR products were sequenced in direct mode using BigDye Terminator Kit and run in automate capillary sequencer according to the manufacturer's instruction. GenBank database, ClustalX, GeneDoc and MEGA softwares were used for phylogenetic analysis.

RESULTS

Discovery of porcine kobuvirus

The initial goal of the study was the detection and molecular epidemiological analysis of porcine caliciviruses in faecal samples of domestic pig by RT-PCR method. Faecal samples were collected from sixty healthy piglets (*Sus scrofa domestica*) in February 2007 from a farm in eastern Hungary. Using the screening primer pair p289/p290 designed for the RNA-dependent RNA polymerase gene region of caliciviruses, sapovirus was detected in 3.3%, while consistently a ~1100 nt long non-specific PCR product for caliciviruses were detected in 35% of the cases. These sequences (swine/S-1-HUN/2007, EU787450) were related to 3C^{pro}/3D^{pol} gene regions of bovine (U-1 virus) and human (Aichi virus) kobuviruses. The complete genome of porcine kobuvirus (S-1-HUN) is 8210 nt long and the coding region is 7467 nt (2488 aa) long. The first 108 nt of the 5'UTR (576 nt) and the secondary RNA structure of the S-1-HUN are very similar to the U-1 and Aichi viruses, but the further part of the 5'UTR containing the IRES is basically different. The S-1-HUN has a *hepacivirus/pestivirus-like* (HP), type IV IRES having every typical motif identified. The 3'UTR region (167 nt) did not have any sequence similarity. L-protein was encoded by the genome. The L, P1, P2 and P3 aa identities are 23/34%, 53/61%, 59/68% and 63/71% to Aichi and bovine kobuviruses, respectively. The largest differences are in the VP1 region within the S-1-HUN P1. The **HWAL** and **NCTHFV** conservative aa motifs are found in 2A from P2 proteins. A sequentially repeating (tandem repeat) 30 aa long motif **AANRVAESIETTAS(/T)V(/A)VREADLARSTLNISM** can be found in 2B (viroporine?). The conserved helicase **GPPGTGKS** aa motif can be found in 2C, while the **GMCGA** (3C) and **KDEL**, **YGDD**, **FLKR**, **GNPSG** (3D) motifs can be seen in P3 proteins. The genome of S-1HUN can be characterized with high G+C (52.4%) ratio.

Detection of porcine kobuvirus with the help of a novel, general kobuvirus screening primer

Based on the 3D gene region of Aichi, U-1 and S-1-HUN kobuviruses a universal screening primer pair (UNIV-kobu-R/UNIV-kobu-F) was designed. The 60 swine faecal samples of healthy animals were retested and 65% of the samples were positive for porcine kobuvirus: in each (100%) of the animals under the age of 10 days, in 93.3% of the 3-week-old animals, in 20% of the 3-month-old animals and in 46.7% of the 6-month-old animals.

Follow-up of porcine kobuvirus infection and virus evolution

The endemic spread and *in vivo* evolution of porcine kobuvirus were investigated in the pig farm where the prototype S-1-HUN virus was originally detected. 21 months after the first sampling 60 faecal/serum sample pairs were retested by RT-PCR method. The porcine kobuvirus RNA was identified in 32 (53.3%) faecal samples, while the virus was found even in serum samples in 16 (26.6%) cases. Nine sample pairs (faeces/serum) were found where the porcine kobuvirus RNA was detected both in faeces and serum, too. The complete genome sequence of one kobuvirus (kobuvirus/swine/K-30-HUN/2008/Hungary; GQ249161) from faecal samples collected during the second sampling was determined and was compared to prototype S-1-HUN virus strain. The highest and lowest nt mutation changes were found in L and 3D regions respectively, while the highest aa mutation changes were in 2C, 2A and VP3 regions. Amino acid changes in VP1, 3A and 3B regions were not experienced.

The first European detection of bovine kobuvirus

Thirty-two faecal samples were collected from cattle in a farm in February 2002. The bovine kobuvirus (kobuvirus/bovine/Aba-Z20/2002/Hungary; FJ225406) was firstly identified in Europe by RT-PCR method using the UNIV-kobu primers in two faecal samples (6.25%) from the 1-7.6-years-old age group.

Characterization and complete genome determination of kobuvirus from sheep

Kobuvirus could be detected by RT-PCR method with the help of UNIV-kobu primers in five of eight (62.5%) faecal samples collected from ovines (*Ovis aries*) in March 2009. The complete genome of ovine kobuvirus (kobuvirus/sheep/TB3-HUN/2009/Hungary; GU245693) is 8378 nt in length and the coding region is 7404 nt (2468 aa) long. The secondary RNA structure of 5'UTR of TB3-HUN (797 nt) is similar to the 5'UTR of the bovine kobuvirus (type V IRES). The 3'UTR region (174 nt) showed sequence similarity to bovine kobuvirus. The L protein is encoded by the genome. The identity of the L, P1, P2 and P3 proteins are 28/54/36%, 48/81/60%, 61/88/67% and 62/92/71% compared to Aichi, bovine and porcine kobuviruses, respectively. The possible cleavage sites of the polyprotein – except 3B/3C borders – were the same as in U-1 virus. The detected ovine kobuviruses were in close phylogenetical relationship with bovine kobuviruses.

Detection of Aichi virus (human kobuvirus) in Hungary

Aichi virus (HUN298/2000/HUN; FJ225407) was detected in one case (1.5%) by RT-PCR using the UNIV-kobu primers out of 65 stool samples which were collected from children under the age of 12 tested negative for bacterial pathogens, rotavirus, adenovirus and calicivirus. The stool sample derived from an out-patient girl who has both upper respiratory and enteric (gastroenteritis) symptoms, too. The HUN298/2000/HUN virus showed 100% aa identity to Aichi virus (BAY-1-03-DEU, AY747174) from Germany, which is classified into Aichi viruses genotype A.

Identification of novel picornavirus from domesticated common quail faecal sample using the UNIV-kobu primers

A strong, specific PCR product was obtained by RT-PCR method using UNIV-kobu primers from a faecal sample of domesticated common quail (*Coturnix coturnix*) in June 2010. The nt sequence showed similarity to the 3D regions of human enterovirus 99 (EF015012) and simian picornavirus 17

(YP_001718553). The complete genome of quail picornavirus (QPV1-HUN/2010, JN674502) is 8159 nt in length (coding region is 7449 nt, 2482 aa). The 5'UTR (494 nt) of the QPV-1 can be described with type IV IRES. Major structural differences can be seen in the IRES, for example a long apical part of domain III. In this place the IRES contains a 20 nt long conservative motif ("8"-like secondary structure), which was found in four more picornaviruses like seal picornavirus (EU142040), pigeon picornavirus B (FR727144), duck hepatitis virus 1 (DQ249299) and turkey hepatitis virus (HQ189775). The 3'UTR (216 nt) of QPV-1 showed partial sequence similarity to 3'UTR of duck sapelovirus sequence. The L protein (with 390 aa, possibly complex function) is encoded by QPV-1 genome, which is rich in cysteine (9.5%) and in which a 34 aa long repetitive motif can be found. The 2A region (11 aa) is excessively short. The P1 (2571 nt, 857 aa), P2 (1374 nt, 458 aa) and P3 (2334 nt, 777 aa) regions demonstrated 43%, 39% and 47% aa identity to avian sapelovirus (formerly name as duck picornavirus, AY563023). The possible cleavage sites of the polyprotein – except the VP4/VP2 and VP2/VP3 borders – were at Q/G borders. The quail picornavirus showed phylogenetic relationship with the almost simultaneously discovered pigeon picornavirus A and B viruses (unclassified picornaviruses) and with sapeloviruses (genus *Sapelovirus*). Nine months later, using the UNIV-kobu primers the quail picornavirus was successfully detected in 30% of the quail faecal samples collected from the same farm. The quail picornavirus RNA could not be detected from the surface of commercially available quail eggs by RT-PCR method.

The first European detection of Enterovirus C109 (EV-C109)

We managed to get a strong specific PCR product with the help of the UNIV-kobu and later EV-C109 specific primers in one case (1.1%) out of 92 nasopharyngeal aspirates collected from hospitalized children with respiratory syndromes in 2005/2006 and 2006/2007 respiratory seasons. The 2.5-year-old boy was treated with bronchitis and pneumonia in January 2007. The partial sequence showed 85% nt and 94% aa identity to coxsackievirus A17 (G12, AF499639) nt

and enterovirus C104 (AGA18499) aa sequences respectively. Comparing the novel prototype enterovirus EV-C109 (NICA-08-4327, GQ965517) sequence, which has been available since 2010, it showed 96% aa identity. The complete genome (7354 nt) of the virus (L87/HUN/2007, JN900470) was determined. P1, P2 and P3 regions have 98%, 98% and 97% aa identity compared to the P1, P2 and P3 regions of prototype EV-C109 strain, respectively. The recombination pattern of the L87 virus coincides with EV-C109: EV-A-like nucleotide sequence in 5'UTR region and EV-C-like nucleotide sequence in the coding and 3'UTR regions.

DISCUSSION AND NOVEL FINDINGS

The non-enveloped picornaviruses with positive, single-stranded RNA (+ssRNA) genome are important human and animal viral pathogens and the most numerous known viral family. The family *Picornaviridae* currently consists of 17 genera and 37 virus species. The number of newly discovered picornaviruses is rapidly increasing; nowadays a lot of new picornaviruses are waiting for taxonomic classification, several of them potentially being the first members of a new genus. Besides, changes affect the basic definition of picornaviruses too. According to the original definition, all picornaviruses have an ORF, but a novel picornavirus (calicivirus A, genus *Dicipivirus*) with an intergenic region containing IRES between the P1 and P2 regions was discovered in 2012 from dogs' faecal samples.

Novel members of *Kobuvirus*, *Enterovirus* and currently unclassified novel picornavirus species, from samples originated from humans and animals, have been reported in this thesis. In the beginning, there were accidental and methodological reasons for the successful identification of viruses containing positive, single-stranded RNA genome. We were working with primers specified for genome parts of viruses but it turned out during the usage of these primers that they work well with unknown viruses, too. This recognition led us to identify novel unknown viruses. Firstly, using the general p289/290 primer pairs for caliciviruses, the porcine kobuvirus was serendipitously detected. Secondly using the general UNIV-kobu-R/UNIV-kobu-F primer pairs for kobuviruses the human Aichi, bovine, ovine and porcine kobuviruses were also detected. Furthermore, the novel quail picornavirus and even a novel human respiratory viral pathogen (EV-C109) were successfully identified with the help of UNIV-kobu-R/UNIV-kobu-F primer pairs too.

The following new findings were achieved in the field of picornaviruses:

- A novel picornavirus species was serendipitously discovered from swine faecal samples by general calicivirus primers (p289/p290). The porcine kobuvirus is classified into genus *Kobuvirus*, which was approved by the ICTV as a new virus species as *Aichivirus C* in March 2013.
- The complete nucleotide sequence of the porcine kobuvirus was determined.
- The untranslated regions (UTRs) of S-1-HUN were analyzed and the secondary RNA structure of the internal ribosome entry site (IRES) of the S-1-HUN was also determined (type IV IRES).
- The coding region of S-1-HUN was analyzed and compared to both prototype Aichi and U-1 kobuvirus sequences. A sequence repeat (tandem repeat) was identified in 2B region.
- The presence of porcine kobuvirus was examined by molecular biological methods from faecal samples of swine in different age groups; frequent and endemic presence of porcine kobuvirus was detected in the investigated pig farm.
- The presence of porcine kobuvirus was confirmed in serum samples herewith the possibility of viraemia was proposed during the kobuvirus infection.
- The *in vivo* evolutionary analysis of porcine kobuvirus was performed from faecal sample (K-30-HUN) collected from the same farm about 2 years later and was compared to the complete genome of the prototype (S-1-HUN) strain. The nucleotide/amino acid changes and frequencies of each gene regions were determined.
- Bovine kobuvirus (*Aichivirus B*) was firstly detected from cattle faecal samples in Europe.
- Kobuvirus (*Aichivirus B*) was also firstly detected from ovine faecal samples. The complete genome of the virus (TB3-HUN) was determined.
- Aichi (human) kobuvirus (*Aichivirus A*) was firstly detected from stool sample originated from children in Hungary.

- A novel picornavirus – quail picornavirus – was identified from quail faecal sample, the complete genome of which was determined and characterized.
- A novel recombinant enterovirus – EV-C109 – was firstly discovered from children nasopharyngeal aspirates under the age of ten suffering respiratory diseases in Europe.
- Molecular biological/genetic approaches were successfully adapted and developed for detection and complete genome determination of picornaviruses.

PUBLICATIONS SERVED AS BASIS FOR THE THESIS

Reuter G., Boldizsár Á., Kiss I., **Pankovics P.**: Candidate new species of kobuvirus in porcine hosts. *Emerg. Infect. Dis.* 2008. 12, 1968-1970. (IF: 6,449)

Reuter G., Boldizsár Á., **Pankovics P.**: Complete nucleotide and amino acid sequences and genetic organization of porcine kobuvirus, a member of a new species in the genus, *Kobuvirus* family *Picornaviridae*. *Arch. Virol.* 2009. 154, 101-108. (IF: 1,909)

Reuter G., Boldizsár Á., Kiss I., Kecskeméti S., **Pankovics P.**: Sertés kobuvírus: egy új vírushaj leírása a *Picornaviridae* család *Kobuvirus* nemzetségében. *Magyar Állatorvosok Lapja*, 2009. 131, 459-461. (IF: 0,2)

Reuter G., Boldizsár Á., Papp G., **Pankovics P.**: Detection of Aichi virus shedding in a child with enteric and extraintestinal symptoms in Hungary. *Arch. Virol.* 2009. 154, 1529-1532. (IF: 1,909)

Reuter G., Kecskeméti S., **Pankovics P.**: Evolution of porcine kobuvirus infection, Hungary. *Emerg. Infect. Dis.* 2010. 16, 696-698. (IF: 6,859)

Reuter G., Boros Á., **Pankovics P.**, Egyed L.: Kobuvirus in domestic sheep, Hungary. *Emerg. Infect. Dis.* 2010. 16, 869-870. (IF: 6,859)

Reuter G., Boros A., **Pankovics P.**: Kobuviruses – a comprehensive review. *Rev. Med. Virol.* 2011. 21, 32-41. (IF: 7,2)

Pankovics P., Boros A., Reuter G.: Novel picornavirus in domesticated common quail (*Coturnix coturnix*). *Arch. Virol.* 2012. 157, 525-530. (IF: 2,03)

Pankovics P., Boros A., Szabó H., Székely G., Gyurkovits K., Reuter G.: Human enterovirus 109 (EV109) in acute paediatric respiratory disease in Hungary. *Acta Microbiol. Immunol. Hung.* 2012. 59, 285-290. (IF: 0,646)

Impact factor of publications served as a base of the thesis: 34.061

Total impact factor: 88.1

Number of independent citations: 296