Genetic characterization of a novel duck-origin picornavirus with six 2A proteins

Xiaoyan Wang, Ning Liu, Fumin Wang, Kang Ning, Yanbo Li and Dabing Zhang

Key Laboratory of Animal Epidemiology and Zoonosis of Ministry of Agriculture, College of Veterinary Medicine, China Agricultural University, 2 Yuanmingyuan West Road, Haidian, Beijing 100193, PR China

A novel virus was detected from diseased ducks and completely determined. The virus was shown to have a picornavirus-like genome layout. Interestingly, the genome contained a total of up to six 2As, including four 2As (2A1–2A4) each having an NPGP motif, an AIG1-like 2A5, and a parechovirus-like 2A6. The 5’UTR was predicted to possess a hepacivirus/pestivirus-like internal ribosome entry site (IRES). However, the subdomain Ille consisted of a 3 nt stem and five unpaired bases, distinct from those found in all other HP-like IRESs. The virus was most closely related to duck hepatitis A virus, with amino acid identities of 37.7%, 39% and 43.7% in the P1, P2 and P3 regions, respectively. Based on these investigations, together with phylogenetic analyses, the virus could be considered as the founding member of a novel picornavirus genus that we tentatively named ‘Aalivirus’, with ‘Aalivirus A’ as the type species.

Picornaviruses are small non-enveloped viruses with single-stranded, positive-sense, polyadenylated RNA genomes. A typical picornavirus genome contains a single long ORF encoding a large protein precursor (polypeptide), which is flanked by 5’ and 3’ untranslated regions (UTRs). The polypeptide is processed co- and post-translationally into 10–14 mature proteins (Fichtner et al., 2013; Knowles et al., 2012).


Duck hepatitis A virus (DHAV) in the genus Avihepatovirus is an important picornavirus that can cause a highly lethal, rapidly spreading infection of young ducklings, and is therefore of significant economic importance in regions with intensive duck production (Woolcock, 2003). On the basis of antigenic and genetic relatedness, three types of DHAV have been identified (DHAV-1–3) (Fu et al., 2008; Kim et al., 2007; Tseng & Tsai, 2007). More recently, a novel picornavirus of turkeys, named avisivirus A (AsV-A), was shown to be related to DHAV (Boros et al., 2013b; Ng et al., 2013). Here, we report the genetic characterization of another novel picornavirus related to DHAV from ducks.

In October 2012, a severe disease occurred in a commercial Pekin duck flock in China, resulting in a morbidity of about 40% in 4- to 5-week-old ducks. Eight intestinal samples collected from 5-week-old diseased ducks were submitted to our laboratory for virus detection. Total RNAs were extracted from the sample using an RNeasy Mini kit (Qiagen). The samples were negative for DHAV using the reverse transcription (RT)-PCR method described by Fu et al. (2008). Application of the primers reported by Lau et al. (2011) in the RT-PCR system produced fragments of approximately 530 bp from two samples. Since the sequences generated from the amplicons were identical, a 486 bp sequence (excluding primer sequences) from one positive sample (designated GL/12) was thus used for further analysis. The sequence was identified by BLASTX search in GenBank (http://blast.ncbi.nlm.nih.gov/Blast.cgi) as encoding a picornavirus-like RNA-dependent RNA-polymerase, which showed a maximum of 51–54% amino acid identity with DHAVs. Pairwise sequence identity and phylogenetic analyses based on amino acid sequences deduced from the 486 nt 3Dpol region indicated that GL/12 was most closely related to, but distinct from, DHAVs (data not shown), suggesting the presence of a novel picornavirus in ducks.

The GenBank/EMBL/DDBJ accession number for the study sequence is KJ000696.

Supplementary material is available with the online version of this paper.
Attempts to propagate the virus in embryonated chicken and duck eggs and several cell lines were all unsuccessful, so the positive sample GL/12 was subjected to full-length sequencing. Most of the genomic sequence of GL/12 was amplified by RT-PCR using the process of primer walking. Primers were designed over conserved regions of DHAVs, AsV-As, human parechoviruses and Ljungan viruses as well as GL/12-specific sequences obtained, using Primer Premier 5.0 (Premier Biosoft International; Table S1 available in the online Supplementary Material). The 5' and 3' ends of the genome were acquired by 5' and 3' rapid amplification of cDNA ends (RACE) strategies (Sambrook & Russell, 2001) (RT-PCR and 5' RACE conditions, cloning and sequencing are shown in the online Supplementary Material). The initial genome sequence was confirmed by designing additional primers (Table S1) to produce seven overlapping RT-PCR products spanning the entire genome with special attention to the 5'UTR and 2A regions.

Sequence analysis using DNAMAN 5.2.2 (Lynnon) revealed that the GL/12 genome comprised 8958 nt, excluding the poly(A) tail. The DNA G+C content (43.2 %) was identical to that of pasivirus A, and similar to those of DHAVs (43.1–43.7 %), cosavirus A (43.8 %), eel picornavirus 1 (43.8 %) and aquamavirus A (AqV-A; 43.9 %) (Table S2). The polyadenylated genomic RNA contained a large ORF (nt 605–8647) of 8043 nt, encoding a putative polyprotein precursor of 2680 aa, which was preceded by a 5'UTR of 311 nt (Fig. 1). The first seven nucleotides at the 5'-end (UUGAAAG) of the GL/12 UTR and 2A regions.

Based on amino acid alignment with DHAVs and AsV-As as well as the NetPicoRNA predictions (Blom et al., 1996), the GL/12 polyprotein was shown to possess a picornavirus-like organization: P1 (VP0-VP3-VP1)-P2 (2A-2B-2C)-P3 (3A-3B-3C-3D) (Figs 1, S1). The P2 region (3756 nt; 1252 aa) showed a significantly longer than P1 (2058 nt; 686 aa) and P3 (2229 nt; 742 aa) and also the largest among the known picornaviruses which range from 459 aa in Seneca Valley virus (Hales et al., 2008) to 1135 aa in melegrivirus A (Honkavuori et al., 2011). The increased length of GL/12 P2 compared with those of other picornaviruses was largely attributed to the presence of an unusually long 2A region (2175 nt; 725 aa) (Figs 1, S1).

Examination of the 2A region revealed that, surprisingly, it contained four canonical cleavage sites: DxE5NPG(705)P, DxE5NPG(988)P, DxE5NPG(1119)P, and DxE5NPG(1288)P (Ryan & Flint, 1997). Proteolytic cleavage at these sites would therefore release four NPGP-containing 2A proteins (a 19 aa 2A1, a 133 aa 2A2, a 150 aa 2A3 and a 131 aa 2A4) and a 292 aa sequence. Based on alignment with DHAVs, the 292 aa sequence may be cleaved at Q(1288)H to generate an additional two distinct 2A products, a 169 aa 2A5 and a 123 aa 2A6. These findings suggested that GL/12 may possess a total of up to six 2A proteins (Figs 1, S1).

To investigate the relationship of GL/12 with other picornaviruses, pairwise comparisons were undertaken based on the amino acid sequences of the P1, P2 and P3 regions of GL/12 and the representative members of the 17 official and nine proposed genera as well as 13 unassigned picornaviruses (Table S2). Pairwise comparisons were performed using CLUSTALW [http://www.genome.jp/tools/clustalw/], with the Gonnet matrix as the comparison scoring table. GL/12 was shown to be most closely related to DHAV types, with average amino acid identities of 37.7 %, 39 % and 43.7 % in the P1, P2 and P3 regions, respectively. According to the picornavirus genus classification rules defined by the ICTV (http://www.picornastudygroup.com/definitions/genus_definition.htm), members of a genus should share >40 %, >40 % and >50 % amino acid identities in the P1, P2 and P3 regions, respectively. GL/12 was therefore identified as a member of a novel genus most closely related to the genus Avihepatovirus. Phylogenetic analyses based on amino acid sequences from the P1, 2C and 3CD regions using

**Fig. 1.** Predicted genome organization of duck picornavirus GL/12. The nucleotide (for UTRs) and amino acid (for each protein) lengths and the predicted cleavage sites are shown.
MEGA5.0 (Tamura et al., 2011) also demonstrated that GL/12 was most related to but highly divergent from DHAVs. The close phylogenetic relationship between GL/12 and AsV-As in the proposed genus ‘Aalivirus’ (Avihepatovirus/Avisivirus-like virus), with ‘Aalivirus A’ (AalV-A) as the type species.

Analysis of the capsid regions revealed that AalV-A contained a potential myristylation sequence (GxxxxS/T) at positions 39–43. This location resembled those of DHAV at positions 31–35 (Ding & Zhang, 2007), canine picornavirus 1 at positions 50–54 (Woo et al., 2012) and feline picornavirus 1 at positions 51–55 (Lau et al., 2012). Whether AalV-A encodes a short leader protein or not remains to be confirmed by analysis of VP0 protein purified from viral particles. Similar to the close relatives avihepatoviruses and avisisviruses, VP0 of AalV-A was probably not cleaved into VP4 and VP2. Using the conserved domain database (CDD) search (Marchler-Bauer et al., 2011), the P1 polypeptide contained Rhv-like domains (pfam entry, cd00205) between residues 92–260 and 322–484. For all three capsid proteins, AalV-A shared the highest amino acid identity values with DHAV, but the levels never exceeded 40% (Table S3).

In terms of length and motif, the 2A1 of AalV-A was comparable to the aphthovirus-like 2A1 of DHAV. Interestingly, high levels of 2A1 sequence identity (69–87%) were found with the aphthovirus-like 2As of porcine teschovirus 1 (PTV-1), hunnivirus A, equine rhinitis B virus and foot-and-mouth disease virus, significantly higher than those (26%) with the similar aphthovirus-like 2A1 of DHAV (Table S3; Fig. S2a). Although the 2A5 possessed only 21% sequence identity to the DHAV 2A2 (Table S3), five characteristic sequence motifs (G1–G5) as well as the switch I and switch II regions, indicative of the GTP binding site of Ras_like_GTPase superfamily of small guanosine triphosphatases (GTPases) (CDD-ID: cl17170) were also identified in the 2A5 protein of AalV-A (Fig. S2b). Thus, the AalV-A 2A5 was related to the AIG1-like 2A2 of DHAVs. As for 2A6, the highest amino acid identity value was observed with the parechovirus-like 2As of DHAV and AsV-A (34%) (Table S3). Based on alignment, the H-box/NC-motifs (Hughes & Stanway, 2000) were detected in the AalV-A 2A6 protein (Fig. S2c), demonstrating that 2A6 belonged to the aphthovirus-like 2A of DHAVs (9%). Using BLASTP searches, the 2A2, 2A3 and 2A4 proteins showed no conserved domains. Using BLASTP searches, the 2A2, 2A3 and 2A4 proteins exhibited little similarity in sequence to 2A proteins of all other picornaviruses.

For most nonstructural proteins (except 2A and 3B), AalV-A shared the highest amino acid identity values with DHAV (Table S3). 3B represented an exception: in this region, AalV-A possessed low sequence identity with DHAV (9%) but showed the highest identity with mosavirus A (27%). The highest amino acid sequence identities were found in 2C (48%), 3D (47%) and 3C (45%) compared with the selected picornviruses. Sequence alignment revealed that nonstructural proteins 2C, 3C and 3D of AalV-A contained the characteristic motifs conserved in picornaviruses. These included motifs typical for helicase (GEKGSKS1740–1747 and DDLGQ1787–1791), cysteine protease (GSGC1818–2189) and RNA polymerase (KDELK2375–2379, DFKKFD2449–2454, GGMCGS8PCTTVLNNT2502–2516, YGDD2542–2545 and FLKR2594–2597).

Prediction of the secondary structure of the 3’-proximal portion of the AalV-A 5’UTR using Mfold (Zuker, 2003) revealed that the region from nt 338 to nt 608 showed structural similarity to hepacivirus/pestivirus (HP)-like type IV IRES with stem–loop domains II and III (Hellen & de Breyne, 2007). As shown in Fig. 3a, the AalV-A domain II contained the E-loop GAA363–365/AGUA379–382 motifs as seen in most of HP-like IRES elements, while domain III included a series of signature elements characteristic of HP-like IRES, e.g. subdomains IIId, IIIe and IIIf. Interestingly, the AalV-A domain IIIe consisted of a 3 nt stem (5’-CUC/5’-GAG) and five unpaired bases (GAUAA), distinct from those of all other HP-like IRESs reported to date (Boros et al., 2013a; Hellen & de Breyne, 2007; Kapoor et al., 2008; Kofstad & Jonassen, 2011; Lau et al., 2011, 2012; Pan et al., 2012; Pankovics et al., 2012; Reuter et al., 2009). Additionally, the sequence to form the proximal half of pseudoknot stem 2 (PK2) was different from the consensus sequences of all three groups (A–C) of HP-like IRESs as summarized by Hellen & de Breyne (2007). These findings suggested that the AalV-A IRES may represent a distinct group of HP-like IRESs.

Detailed examination revealed that AalV-A possessed a unique mixture of HP-like IRES characteristics (Fig. 3a). Domain II in AalV-A contained a stretch of 24 nt (GUGA-AACGGAUUACCAGGUAGUAGC), identical to those found in DHAVs (Hellen & de Breyne, 2007; Pan et al., 2012), which formed the apical region of domain II. The predicted apical half of domain III was most similar to that of AqV-A (Kapoor et al., 2008). The sequences of the proximal half of PK1 (5’-ACUAG/5’-CUAUG) and the helix III1 (5’-CGA/5’-ACG) were identical to those of classical swine fever virus (CSFV) (Hellen & de Breyne, 2007). Similar to those found in group C HP-like IRESs (Hellen & de Breyne, 2007), PK1 of AalV-A IRES consisted of 8 nt and 3 nt helices linked by an asymmetrical bulge. Like CSFV, AqV-A, feline picornavirus 1 and bat picornaviruses 1 and 2 (Hellen & de Breyne, 2007), the subdomain IIId of AalV-A IRES contained an unpaired GGGG motif within the terminal loop. The length of PK2 (6 nt) was identical to those of hepatitis C virus and hepatitis B virus in groups A IRES, whereas the length of linker (6 nt) was identical to that of bovine viral diarrhea virus in group C (Hellen & de Breyne, 2007). As for the length of spacer (7 nt) between the PK2 and the start codon, it was closest to those of PTV-1 and simian sapelovirus 3 in group B HP-like IRES (Hellen & de Breyne, 2007). Additionally, AalV-A also possessed an ‘8-like’ 20 nt-long
Fig. 2. Phylogenetic relationship of duck picornavirus GL/12 with representatives from 17 official and nine proposed picornavirus genera and 13 unassigned picornaviruses. The trees were constructed based on the amino acid sequences of P1 (a), 2C (b) and 3CD (c), using the neighbour-joining method with the Poisson model and 1000 bootstrap replications. (●) Strain determined in this study. Bars, 0.1 substitutions per amino acid position.
conserved structure before domain II as seen in DHAV and AqV-A (Kapoor et al., 2008; Pan et al., 2012).

The 3'UTR of AalV-A was similar in length and predicted secondary structure to DHAV-1 (Fig. 3b; Boros et al., 2012). By analogy to DHAV-1, GV-A, passerivirus A, a newly characterized chicken picornavirus and some kobuviruses (Boros et al., 2012; Bullman et al., 2014), the 3'UTR of AalV-A also contained a conserved 48 nt 'barbell-like' structure, with a 9 + 6 nt long identical nucleotide motif in the lower loop region and a polypyrimidine tract at the upper loop region. Compared to other picornviruses, AalV-A possessed a relatively long polypyrimidine tract (UUUUCUUUU).

Taken together, we identified a novel duck picornavirus that could be considered as the founding member of a novel genus in the family Picornaviridae. While the virus exhibited a picornavirus-like genome layout, it had two unique features. One was that the genome possessed a total of up to six 2A regions, and another was that the virus contained a distinct HP-like IRES. Further studies are needed concerning this newly identified picornavirus, especially its pathogenicity and the role of the six different 2A proteins in viral replication.
Novel picornavirus in ducks

Fig. 3. (a) Predicted secondary structure of the IRES of AalV-A GL/12, the complete genome sequence (KJ000696) of which is given in the online Supplementary Material. Domain numbers and pseudoknot stem elements are labelled according to Hellen & de Breyne (2007). The E-loop motifs in domain II are indicated by grey shading. Dotted boxes indicate nucleotides identical to DHAVs. The start codon is underlined. (b) Predicted secondary 3'UTR structure of AalV-A GL/12. Dotted box indicates the conserved ‘barbell-like’ structure. Grey boxes indicate the 9 + 6 nt long identical nucleotide motif (Boros et al., 2012). The stop codon is underlined.

Acknowledgements

This work was supported by the China Agriculture Research System (CARS-43) and Special Fund for Agro-scientific Research in the Public Interest (201003012).

References


